

Bootstrap inference for conditional risk measures

Citation for published version (APA):

Heinemann, A. M. (2019). *Bootstrap inference for conditional risk measures*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20190620ah>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190620ah](https://doi.org/10.26481/dis.20190620ah)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Bootstrap Inference for Conditional Risk Measures

Alexander Heinemann

© Alexander Heinemann, Maastricht 2019

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by ProefschriftMaken || www.proefschriftmaken.nl

ISBN: 978 94 6380 372 4

Printed in The Netherlands by ProefschriftMaken || www.proefschriftmaken.nl

Bootstrap Inference for Conditional Risk Measures

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday, 20 June, 2019 at 16.00 o'clock

by

Alexander Maria Heinemann

Supervisor

Prof. dr. F.C. Palm

Co-Supervisors

Dr. E.A. Beutner

Dr. S.J.M. Smeekes

Assessment Committee

Prof. dr. A.W. Hecq (Chair)

Dr. N. Bastürk

Prof. dr. G. Cavaliere, University of Bologna, Italy

Prof. dr. C. Francq, CREST, Paris, France

Prof. dr. P.C. Schotman

This research was financially supported by The Netherlands Organisation for Scientific Research (NWO).

In memory of Christa Potetzki

Acknowledgements

*“You cannot connect the dots looking forward,
you can only connect them looking backwards.”*

-Steve Jobs (1955-2011)

Looking backwards, the starting point of my academic journey is a letter from Maastricht University, back in July 2010, congratulating me on the successful admission to the *International Business* program. I soon came to know that my primary study choice *International Business Economics* belongs to *Economics* track. Misled by the ambiguous naming, I retrospectively thank the admission office for not allowing me to switch to *Economics*. Acting by necessity, I switched to *Econometrics and Operations Research* instead – a decision I have never regret ever since. Nine years of intensive studies followed. Having reached the mountain’s top – my PhD defense – I would like to thank several people in particular, who accompanied me along the way.

First of all, I am greatly indebted to my supervisors and mentors Eric Beutner and Stephan Smeekes. It truly has been a privilege to work with and learn from the both of them. I consider myself extremely lucky to have not only one but two supervisors, who care so much about my work and are approachable anytime. Eric, I especially would like to thank you for your encouragement and for the numerous counter-examples invalidating former versions of the proofs. Stephan, I benefited a lot from your constructive remarks and bootstrap insights. In addition, I owe my promoter Franz Palm a debt of gratitude for reading several drafts and providing comments that greatly improved the work presented herein.

Further, I would like to thank the reading committee for taking the time to carefully evaluate and approve my thesis manuscript. In particular, I need to mention Christian Francq; he and Jean-Michel Zakoian gave me a warm welcome at CREST, Paris, where I spent a fantastic time as a visiting researcher. *Merci!* I hope we find the time to collaborate more in the future. A special thanks also goes to my former supervisor and friend Dante Amengual, who guided me through the reviewing process of the *Journal of Applied Econometrics* and who has a great share in publishing my first article. *Muchas gracias!*

Let me turn the attention to my paranympths, Sean and Rasmus, who shared with me the up’s and down’s of the PhD. I cannot put in words how grateful I am for your pure friendship and the daily distractions. Sean, I already miss your infectious laughter and the inspiring voice messages after midnight. Rasmus, my Swedish friend, I highly appreciate your spontaneity and your sense of humor; sharing an office with you for over three years was a terrific experience. I also would like to thank Anne Balter and Farzaneh Rajabi Ghamchi for the wonderful atmosphere in office at the early and the final stage of my PhD.

ACKNOWLEDGEMENTS

A warm word of thanks goes to Yolanda Paulissen and Karin van den Boorn who facilitated my research activity by taking most of the administrative burden upon themselves. *Van haarte bedankt!* I also send my appreciation to my other colleagues, especially those who contributed to the legendary atmosphere on the 4th floor at SBE:

Ines Wilms	Jan van den Brakel	Celia Paquay	Niloufar Yousefmanesh	Jean-Pierre Urbain	Ton Storeken	Stan van Hoesel	Kim van Broekhoven	Sanne van Wetten
Tim Oosterwijk Alexis	Hans de	Graaff	Etienne	János	Flesch	Yuliya	Shapovalova	Alexandra de
Derumigny Thomas van	Aditya	Aradiye	Wijler Rui	Yicong Lin	Henrik	Marina	Friedrich	Gendre Aida Abiad
Gemert	Hanno	Reuvers	Chen Nina	Zaunbrecher	Moritz	Christian	Kerkoffs	Lenard Lieb
Vincent	Dries Vermeulen	Andrey	Kart haus	Buchem Andre	Berger	Johannes	Schniffels	Juan Palacios
Keuzen	Kateshov Jan Lohmeyer	Alain	Hecq	Denis de	Crumbrughe	Alexander	Grigoriev	Dewi Peerlings
Roland	Simon	Reese	Sébastien Vries Niels	Veerle	Timmermans	Inka Eber	hardt Tim	Nina Scherl
Vinceze	Shash	Khare	Mourmans Lidwien Sol	Luca	Magaritella	Reissner	Tom Stolp	Caterina Schiavoni
Li Sun	Anna	Zseleva	Julian	Golak	Diogo	Geraldes	Benoit Duvocelle	Marc Schröder
Thijs Kamma Adam Jassen Timma Bressgott Hannah Schmitt Elisa Voisin Nora Pankratz Verena Jung Alexander Grigoriev Rui Jorge De Almeida e Santos Nogueira								

I express my gratitude to Peter Thesling, Frank Bosserhoff and Florentijn Hogerwerf for a *gezellige tijd* and José Jimenez, Andreas Stegmann and Diego Astorga for the nice company and a great student time. Outside the academic circle I can always count on my lifelong friends Fabian Ziehms and Matthias Schöfeld. You are the best! I also cannot get around to thank Alix Greiber, who guided me back to Maastricht. The *Maashattan* experience would not have been the same without Manos. Nine generations of Manos are simply too many to mention every single person, but I will never forget the memorable handball tournaments and member weekends.

Of course, I would like to acknowledge my family, the anchor in my life. *Mama*, thank you for the constant support over all these years and for always believing in me. *Papa*, my sincere apologies: you tried so hard to talk me out of mathematics, however *ein Apfel fällt nicht weit vom Stamm*.¹ Special credits also go to my sisters Isabel, Kathi and Sarah. And I thank my penfriend Fabian, my princess Julia and my godchild Emil for being so adorable. Saving the best for last: Suus, I am truly blessed to have you in my life. Thank you for your patience and all the distracting smiles during stressful times.

Alexander Heinemann
Maastricht, January 2019

¹German proverb equivalent to *Like father, like son*.

Contents

“I was still a couple of miles above the clouds when it broke, and with such violence I fell to the ground that I found myself stunned, and in a hole nine fathoms under the grass, when I recovered, hardly knowing how to get out again. Looking down, I observed that I had on a pair of boots with exceptionally sturdy straps. Grasping them firmly, I pulled with all my might. Soon I had hoist myself to the top and stepped out on terra firma without further ado.”

-Rudolf Erich Raspe (1736-1794)

Contents	xi
1 Introduction	1
1.1 Risk Measures	3
1.2 Classical Asymptotic vs. Bootstrap Inference	4
1.3 Temporal Dependence	7
1.4 Contribution and Outline of the Thesis	8
2 A Justification of Conditional Confidence Intervals	11
2.1 Introduction	12
2.2 General Setup	14
2.2.1 The General Prediction Function	14
2.2.2 Argument of Two Independent Processes	21
2.2.3 Sample-split Estimation	22
2.3 Asymptotic Justification	24
2.3.1 Merging	24
2.3.2 Merging of 2IP and SPL in Probability	26
2.3.3 Interval Construction	29
2.3.4 Interval Construction Under Normality	32
2.4 Prediction Intervals	34
2.5 Conclusion	36
2.A Lemmas and Proofs of Theorems	38
2.A.1 Lemmas	38
2.A.2 Proofs of Theorems.	40
2.B Additional Proofs	46
2.B.1 Proofs of Lemmas	46
2.B.2 Proofs of Corollaries	54
3 A General Framework for Prediction in Time Series Models	57

3.1	Introduction	58
3.2	General Framework	59
3.3	Conditional Mean in an AR(1)	60
3.3.1	Model Description	60
3.3.2	Estimation	61
3.3.3	Mapping	61
3.3.4	Verification of Assumptions	61
3.4	Conditional Variance in a GARCH(1,1)	64
3.4.1	Model Description	64
3.4.2	Estimation	65
3.4.3	Mapping	66
3.4.4	Verification of Assumptions	67
3.5	Conditional Mean in an ARMA(1,1)	72
3.5.1	Model Description	72
3.5.2	Estimation	73
3.5.3	Mapping	74
3.5.4	Verification of Assumptions	75
3.6	Conditional Volatility in a T-GARCH(1,1)	81
3.6.1	Model Description	81
3.6.2	Estimation	82
3.6.3	Mapping	83
3.6.4	Verification of Assumptions	84
3.7	Concluding Remarks	90
4	A Residual Bootstrap for Conditional Value-at-Risk	93
4.1	Introduction	94
4.2	Model	96
4.3	Estimation	97
4.4	Bootstrap	103

4.4.1	Fixed-design Residual Bootstrap	103
4.4.2	Bootstrap Consistency	105
4.4.3	Bootstrap Confidence Intervals for VaR	108
4.5	Numerical Illustration	109
4.5.1	Monte Carlo Experiment	109
4.5.2	Empirical Application	120
4.6	Concluding Remarks	121
4.A	Auxiliary Results and Proofs	123
4.A.1	Non-bootstrap Lemmas	123
4.A.2	Bootstrap Lemmas	148
4.B	Recursive-design Residual Bootstrap	184
5	A Residual Bootstrap for Conditional Expected Shortfall	187
5.1	Introduction	188
5.2	Model	189
5.3	Estimation	190
5.4	Bootstrap	195
5.4.1	Fixed-design Residual Bootstrap	195
5.4.2	Bootstrap Consistency	196
5.4.3	Bootstrap Confidence Intervals for ES	198
5.5	Monte Carlo Experiment	198
5.6	Conclusion	203
5.A	Auxiliary Results and Proofs	205
5.B	Derivation of Analytical Expressions	213
6	Conclusion	217
7	Valorisation	223
	Bibliography	227

Nederlandse Samenvatting	237
Curriculum Vitae	241

Chapter 1

Introduction

*“The advantage of knowing about risks is that
we can change our behavior to avoid them.”*

-Robert F. Engle (1942-...)

Risk is the exposure to some potential loss resulting from uncertainty. With various sources of uncertainty in financial markets, market participants have developed numerous concepts to deal with risk that are collectively known as *risk management*. Financial risk management has tremendously evolved in past decades becoming an increasingly sophisticated practice with central function for businesses. Its key objective is to identify, quantify and manage the risk that a financial institution faces in order to ensure long-term success. While risk taking is a lucrative business for banks and insurance companies in fortunate events, it may be harmful when unfortunate events arise (e.g. the financial crisis 2007/08). Concerned about excessive risk-taking, legislators all over the world enforce minimum standards for financial risk management. In banking such regulatory framework is called Basel III while in the insurance sector similar requirements exist known as Solvency II. Both, Basel III and Solvency II, impose capital requirements, i.e. banks and insurance companies are required to keep a reserve to compensate for the risk they are exposed to. Financial institutions frequently resort to so-called *risk measures* to calculate the capital reserves and to assess the market risks resulting from fluctuations of financial assets.

1.1 Risk Measures

Mathematically, a risk measure is defined as a mapping from the set of all real-valued random variables, denoted by \mathcal{X} , onto the real line. In other words, a risk measure ρ summarizes the risk associated with a random variable X in a single number, which we denote by $\rho(X)$. Artzner, Delbaen, Eber, and Heath (1999) propose a set of desirable properties for risk measures to fulfill:

Axioms. (*Coherent risk measures*)

1. (*Translation invariance*) $\rho(X + \alpha) = \rho(X) - \alpha$ for all $X \in \mathcal{X}$ and all $\alpha \in \mathbb{R}$;
2. (*Subadditivity*) $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$;
3. (*Positive homogeneity*) $\rho(\alpha X) = \alpha \rho(X)$ for all $\alpha \geq 0$;
4. (*Monotonicity*) $\rho(X_2) \leq \rho(X_1)$ for all $X_1, X_2 \in \mathcal{X}$ with $X_1 \leq X_2$.

The first axiom means that adding a risk-free amount α to the initial position X simply decreases the risk measure by α . The property of *subadditivity* states

that the risk associated with diversifying among two positions must be less than or equal to the sum of their individual risks. The axiom of *positive homogeneity* implies that doubling an initial position X , yields a risk, which is twice as large. Last, *monotonicity* entails that if a position outperforms another position in all states of the world, then its risk must be comparably smaller or equal. A measure that satisfies all four axioms is called *coherent*. The simplest example of a risk measure is the standard deviation also known as volatility.

Example 1.1. The volatility $\sigma(X) = (\text{Var}[X])^{\frac{1}{2}}$ is a coherent risk measure.

Often notation is shortened writing σ instead of $\sigma(X)$. Arguably the most popular risk measures in finance are *Value-at-Risk* and *Expected Shortfall*.

Definition 1.1. (*Value-at-Risk*) Given a level $\alpha \in (0, 1)$, the α Value-at-Risk (VaR) is defined by $\text{VaR}_\alpha(X) = -\inf\{x \in \mathbb{R} : \mathbb{P}[X \leq x] \geq \alpha\}$.

In words, the VaR is the largest value such that the probability that the position is less than VaR is α . Given a VaR exceedance has occurred, the Expected Shortfall (ES) is simply the expected value of the position.

Definition 1.2. (*Expected Shortfall*) Given a level $\alpha \in (0, 1)$, the α ES is defined by $\text{ES}_\alpha(X) = \mathbb{E}[X | X < \text{VaR}_\alpha]$.

Whereas ES is a coherent risk measure, VaR does not satisfy (in general) the subadditivity axiom. Nevertheless, it enjoys great popularity among practitioners due to its conceptional clarity.

1.2 Classical Asymptotic vs. Bootstrap Inference

The risk measures described in the previous section are quantities of the risk population and therefore treated as parameters. Such parameters are usually unknown to the researcher and need to be inferred from the data. In statistics, one distinguishes between point estimation and interval estimation. Whereas point estimation yields a single number as an estimate, interval estimation gives a range of values that is likely to contain the parameter. Subsequently, we focus on the volatility measure, albeit the discussion carries easily over to other risk measures. Suppose one observes a sample of data X_1, \dots, X_n , which is assumed to be independent and identically distributed (i.i.d.).¹ Given the data a natural point

¹This assumption is clearly violated for financial data and will be relaxed at a later stage.

estimator for the volatility σ is

$$\hat{\sigma}_n = \left(\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 \right)^{\frac{1}{2}} \quad (1.1)$$

with $\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$. Standardizing (1.1), its finite sample distribution $G_n(x) = \mathbb{P}[\sqrt{n}(\hat{\sigma}_n - \sigma) \leq x]$ is usually unknown. To obtain an interval that contains the parameter σ with pre-specified probability, say 95%, one traditionally relies on asymptotic theory, in which the sample size approaches infinity. Frequently the limiting distribution $G_\infty(x)$ depends on nuisance parameters, which need to be replaced by consistent estimators. For instance, under regulatory conditions, asymptotic theory yields that the $G_n(x)$ approaches a normal distribution with mean zero and some variance ς^2 . Replacing ς by a consistent estimate², say $\hat{\varsigma}_n$, the asymptotic normality implies that the probability that the following interval

$$\left[\hat{\sigma}_n - 1.96 \frac{\hat{\varsigma}_n}{\sqrt{n}}, \hat{\sigma}_n + 1.96 \frac{\hat{\varsigma}_n}{\sqrt{n}} \right] \quad (1.2)$$

contains the parameter σ converges to 95% as $n \rightarrow \infty$.

A powerful alternative to asymptotic theory for performing statistical analysis is the bootstrap. The method's name is derived from the phrase “to pull oneself up by one's bootstrap”, which is widely thought to originate from *The Surprising Adventures of Baron Münchhausen* by Raspe (1785). In the famous tale, the main character pulls himself out of a swamp by his own bootstrap. While this seems physically impossible, the statistical crux of the bootstrap is that the sample –and only the sample– give rise to its own statistical properties. This is achieved by treating the data as if they were the population and then drawing new samples from it, which ought to mimic the statistical properties of the original sample. In that sense the bootstrap method is a simple algorithmic procedure, which gained large popularity among practitioners since its introduction by Efron (1979). The following algorithm illustrates the construction of a bootstrap confidence interval for the volatility parameter σ .

Algorithm 1.1. (*Volatility i.i.d. bootstrap*)

1. Generate a bootstrap sample X_1^*, \dots, X_n^* by randomly drawing with replace-

²Because $\varsigma^2 = \frac{\kappa - \sigma^4}{4\sigma^2}$ with $\kappa = \mathbb{E}[(X - \mathbb{E}[X])^4]$, a natural estimator for ς is given by $\sqrt{\frac{\hat{\kappa}_n - \hat{\sigma}_n^4}{4\hat{\sigma}_n^2}}$, where $\hat{\kappa}_n = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^4$.

ment from X_1, \dots, X_n

2. Calculate $\hat{\sigma}_n^* = \left(\frac{1}{n} \sum_{t=1}^n (X_t^* - \bar{X}_n^*)^2 \right)^{\frac{1}{2}}$ with $\bar{X}_n^* = \frac{1}{n} \sum_{t=1}^n X_t^*$
3. Repeat the previous steps B times and let $\hat{\sigma}_n^{*(b)}$ denote the bootstrap estimator in the b -th iteration. Estimate $G_n^*(x) = \mathbb{P}^*[\sqrt{n}(\hat{\sigma}_n^* - \hat{\sigma}_n) \leq x]$ by

$$G_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}(\hat{\sigma}_n^{*(b)} - \hat{\sigma}_n) \leq x\}},$$

where $\mathbb{1}_{\{A\}}$ is equal to 1 if event A is true and zero otherwise.

The bootstrap quantities are conventionally denoted using an asterisk or star superscript. The number of bootstrap replications B in step 3 involves a trade-off between estimation accuracy and computational time: the larger B , the more accurate the estimate, however the more time the computer needs to perform the calculations. A bootstrap interval for σ analog to (1.2) can be found by selecting the 2.5% and 97.5% quantile of $G_{n,B}^*(x)$, denoted by $G_{n,B}^{*-1}(0.025)$ and $G_{n,B}^{*-1}(0.975)$, and setting

$$\left[\hat{\sigma}_n - \frac{G_{n,B}^{*-1}(0.975)}{\sqrt{n}}, \hat{\sigma}_n + \frac{G_{n,B}^{*-1}(0.025)}{\sqrt{n}} \right].$$

Proving the validity of the bootstrap amounts to show that bootstrap distribution G_n^* and the finite sample distribution G_n are close in some sense.

Definition 1.3. (*Bootstrap consistency*) Let G_n be the finite sample distribution and G_n^* denote the corresponding bootstrap distribution. Given a distance d , the bootstrap method is consistent if $d(G_n, G_n^*) \rightarrow 0$ in probability as $n \rightarrow \infty$.

Common choices for the distance d are the Kolmogorov or the bounded Lipschitz distance. In the presence of a continuous limiting distribution G_∞ , several distances yield equivalent³ definitions and bootstrap consistency follows from verifying that $d(G_n^*, G_\infty) \rightarrow 0$ in probability as $n \rightarrow \infty$.

³Formally, the distances induce the same topology.

1.3 Temporal Dependence

Up to this point, we presumed the data to be i.i.d. and left the time dimension aside. However, most financial data possess a temporal ordering, which frequently reveals unique features of the data. Figure 1.1 plots the log-returns of the S&P500

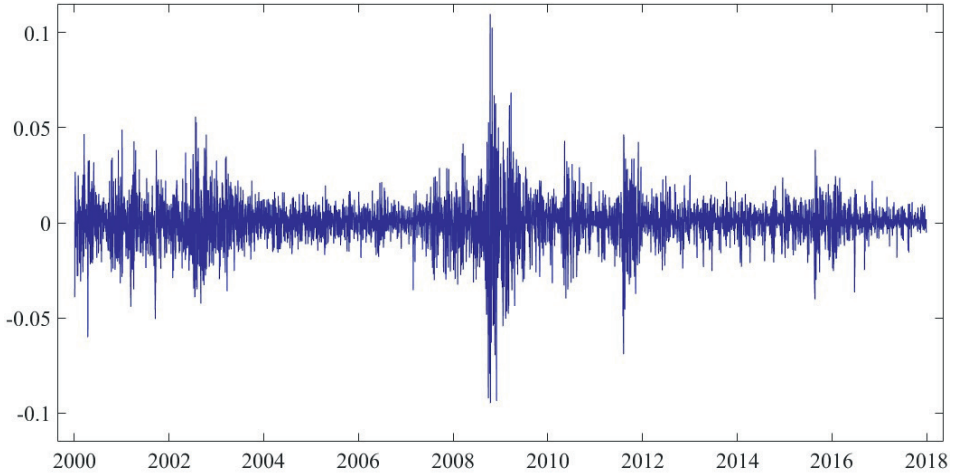


Figure 1.1: Daily Log-return of the S&P500 index for the years 2000 until 2017.

index for the years 2000 until 2017. We observe that “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes” (Mandelbrot, 1963), which is known as *volatility clustering*. To reflect such temporal dependence, we introduce the conditional probability given the past $\mathbb{P}_{t-1}[\cdot] = \mathbb{P}[\cdot | X_{t-1}, X_{t-2}, \dots]$ and denote the conditional mean and conditional variance associated with \mathbb{P}_{t-1} by \mathbb{E}_{t-1} and $\mathbb{V}\text{ar}_{t-1}$, respectively. In correspondence to the unconditional volatility measure σ , its conditional counterpart $\sigma_t = (\mathbb{V}\text{ar}_{t-1}[X_t])^{\frac{1}{2}}$ depends on X_{t-1}, X_{t-2}, \dots and is therefore a stochastic process. Similarly, one obtains conditional analogues for VaR and ES.

Definition 1.4. (*Conditional Value-at-Risk*) Given a level $\alpha \in (0, 1)$, the conditional α VaR is defined by $\text{VaR}_{t,\alpha}(X_t) = -\inf\{x \in \mathbb{R} : \mathbb{P}_{t-1}[X_t \leq x] \geq \alpha\}$.

Definition 1.5. (*Conditional Expected Shortfall*) Given a level $\alpha \in (0, 1)$, the conditional α ES is defined by $\text{ES}_{t,\alpha}(X_t) = \mathbb{E}_{t-1}[X_t | X_t < \text{VaR}_{t,\alpha}]$.

Again, these risk measures are unobserved by the researcher and must be inferred from the data. For this purpose, the stochastic process of the risk measure

is typically parameterized, e.g. $\sigma_t = \sigma(X_{t-1}, X_{t-2}, \dots; \theta)$, and an estimate of the risk measure is obtained by replacing the unknown parameters by estimates.

1.4 Contribution and Outline of the Thesis

Several bootstrap methods have been proposed in the literature to quantify the uncertainty around the risk measures' point estimates. Whereas existing work shows that simulation results are promising, there is, however, no theoretical result underpinning the validity of these methods.

The aim of this thesis is to fill this gap. In particular, we can identify three principal contributions of this thesis. First, we provide a novel, and realistic, justification for commonly constructed intervals around point estimates of conditional objects such as conditional risk measures. This asymptotic justification is of rather theoretical nature and established for a general class of time series models. Second, we propose bootstrap methods to mimic the finite sample distribution of the quasi-maximum-likelihood estimator associated with the conditional VaR and ES. In addition, we give formal proofs of bootstrap consistency, which confirm the validity of those methods. Third, we conduct simulation studies to evaluate the performance of the bootstrap methods in finite samples. Several interval types are investigated and recommendations are made regarding the choice of the interval, which serves as practical guideline for practitioners.

Two challenges arise throughout this thesis. The first issue possesses a fundamental character and stems from the fact that on one hand one must *condition* on the sample as the past informs about the present, yet on the other hand one must allow the data up to now to be treated as *random* to account for estimation uncertainty. The second issue originates from the stochastic nature of the conditional risk measures. Because the latter vary over time and they do not permit a limiting distribution, which further complicates the analysis.

Both challenges are addressed in Chapter 2 in detail. Typically, researchers resort to the unrealistic assumption of two-independent processes to bypass the fundamental issue described above. To avoid this assumption, a solution based on a simple sample-split approach is proposed in this chapter, which requires a much more realistic weak dependence condition instead. Further, to acknowledge that the conditional quantities vary over time, a merging concept is employed, which generalizes the notion of weak convergence.

Chapter 3 verifies the high-level assumptions of the previous chapter for various time series models including autoregressive moving-average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) type models.

In Chapter 4 we propose a fixed-design residual bootstrap method for the two-step estimator of Francq and Zakoïan (2015) associated with the conditional VaR. The bootstrap's consistency is proven under mild assumptions and bootstrap intervals are constructed for the conditional VaR. A large-scale simulation study supports the theoretical results and sheds further light on the bootstrap's performance in finite samples.

Chapter 5 builds on the results of Chapter 4. In this chapter a fixed-design residual bootstrap is investigated for the estimator of the conditional ES. The asymptotic validity of the bootstrap scheme is established and a Monte Carlo experiment is carried out to study the finite sample behavior. Simulations confirm that the method performs adequately in samples of modest size.

Chapter 6 provides a short conclusion of the thesis. Proofs of the theoretical results are collected in appendices at the end of each chapter. Whereas notation is consistent within each chapter, it may differ across chapters. Therefore relevant notation is specified within the chapters.

Chapter 2

A Justification of Conditional Confidence Intervals

To quantify uncertainty around point estimates of conditional objects such as conditional means or variances, parameter uncertainty has to be taken into account. Attempts to incorporate parameter uncertainty are typically based on the unrealistic assumption of observing two independent processes, where one is used for parameter estimation, and the other for conditioning upon. Such unrealistic foundation raises the question whether these intervals are theoretically justified in a realistic setting. This chapter presents an asymptotic justification for this type of intervals that does not require such an unrealistic assumption, but relies on a sample-split approach instead. By showing that our sample-split intervals coincide asymptotically with the standard intervals, we provide a novel, and realistic, justification for confidence intervals of conditional objects. The analysis is carried out for a rich class of time series models.¹

¹This chapter is based on the paper Beutner, Heinemann, and Smeekes (2019b).

2.1 Introduction

One of the open questions in time series is how to quantify uncertainty around point estimates of conditional objects such as conditional means or conditional variances. A fundamental issue arises in the construction of confidence intervals that ought to capture the parameter estimation uncertainty contained in these objects. This fundamental issue stems from the fact that on one hand one must *condition* on the sample as the past informs about the present, yet on the other hand one must allow the data up to now to be treated as *random* to account for estimation uncertainty. The issue is well-recognized in the econometric literature, however in practice confidence intervals are commonly constructed by treating the sample *simultaneously* as fixed and random. Frequently, such approach is motivated by presuming to have two independent processes. Assuming two independent processes with the same stochastic structure, using one for conditioning and one for the estimation of the parameters, bypasses the issue. It is a mathematically convenient assumption as in such case the uncertainty quantification reduces to an ordinary inferential problem. However, practitioners rarely have a replicate, independent of the original series, at hand with the exception of perhaps some experimental settings. As such, the intervals commonly constructed by practitioners lack a satisfactory theoretical justification. Therefore it is the objective of the present work to develop a realistic justification for such confidence intervals around point estimates of conditional objects.

In the literature the fundamental issue described above is encountered in various ways. In the specific case of a first-order autoregressive (AR) process with Gaussian innovations, Phillips (1979) investigates the statistical dependence between the ordinary least squares (OLS) estimator and the endogenous variable conditioned upon. He obtains an Edgeworth-type expansion for the distribution of the conditional mean and, further, studies forecasting, where the fundamental issue equally arises.² Lütkepohl (2005, p. 95) explicitly states a two-independent-processes assumption in connection with vector AR models. He postulates that such assumption is asymptotically equivalent to using only data not conditioned upon for estimation. Ing and Wei (2003) clearly distinguish between *independent-realization* and *same-realization* settings and study the unconditional mean-squared prediction error in the latter for an infinite-order AR process. In a companion paper they also provide a theoretical verification for order selection

²For prediction intervals some solutions have been discussed. We refer to Section 2.4.

criteria for same-realization predictions and stress that it can be misleading to assume that the results for independent-realization settings carry over to those for corresponding same-realization cases (Ing and Wei, 2005). Other studies investigate parameter uncertainty by using resampling methods, that typically mimic a distribution in which the sample, or at least a subsample, is treated as fixed and random at the same time (cf. Pascual, Romo, and Ruiz, 2004, 2006, Pan and Politis, 2016a, 2016b). Aware of this paradox, Kreiss (2016) points out that conditioning on observing specific *in-sample* values affects the parameter estimator, but the effect is often erroneously disregarded. Deviating from the various bootstrap approaches, Hansen (2006) examines parameter uncertainty in interval forecasts in a classical statistical framework. Similar to a general regression framework, he conditions on an arbitrary fixed *out-of-sample* value to avoid the issue. However, conditioning on arbitrary fixed out-of-sample values appears incompatible with the usual setup of dynamics in which we condition on the final value(s) of the sample. Acknowledging the issue while avoiding the two-independent-processes argument bears careful statements as in Francq and Zakoïan (2015) who write in view of this issue “the delta method ... suggests” (p. 162). Similarly, Pesaran (2015) notices that although such intervals “have been discussed in the econometrics literature, the particular assumptions that underlie them are not fully recognized” (p. 389).

This chapter provides a novel, and realistic, justification for commonly constructed confidence intervals around point estimates of conditional objects. Our solution is based on a simple sample-split approach and a weak dependence condition, which allows to partition our sample into two asymptotically independent subsamples. For a rich class of time series models we construct asymptotically valid sample-split intervals, without relying on the assumption of observing two independent processes, and show that these intervals coincide asymptotically with the intervals commonly constructed by practitioners. As will be argued below, an appropriate concept to study conditional confidence intervals is merging, a concept that generalizes weak convergence. To the best of our knowledge, except for Belyaev and Sjöstedt-De Luna (2000), the present work is the only one to study merging in the context of conditional distributions. Moreover, we seem to be the first to employ merging of conditional distributions in time series. By employing this concept we avoid unnatural assumptions such as observing $X_n = x$ (in dynamic models), losing the time index n , and instead explicitly acknowledge that the conditional objects vary over time.

The rest of this chapter is organized as follows. Section 2.2 specifies the general setup and describes the argument of two independent processes as well as our sample-split approach. In Section 2.3 we establish merging among the proposed and the two-independent-processes estimator in probability under mild conditions. Further, we construct asymptotically valid sample-split intervals and show that these coincide asymptotically with the standard intervals. The extension to prediction is discussed in Section 2.4. Section 2.5 concludes. The main proofs are collected in Appendix 2.A, while Appendix 2.B provides additional proofs of intermediate results.

2.2 General Setup

2.2.1 The General Prediction Function

Let $\{X_t\}$ be a real-valued stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. θ denotes a generic parameter vector of length $r \in \mathbb{N}$ and θ_0 the true value, unknown to the researcher.³ Let $\Theta \subseteq \mathbb{R}^r$ be the corresponding parameter space.

Our general setup involves inference on an object that we call the *prediction function*, which is a function of both the process $\{X_t\}$ and of the parameter θ . It represents the random object of interest, and will typically express quantities such as a conditional mean or conditional variance (without conditioning on a specific value) as a function of the sample.

Definition 2.1. The *prediction function* $\psi : \mathbb{R}^\infty \times \Theta \rightarrow \mathbb{R}$ is depending both on the parameter θ and the entire history of the process $\{X_t\}$, such that we can write the prediction of the quantity at time $n + 1$, using data up to time n , as

$$\psi_{n+1} := \psi(X_n, X_{n-1}, \dots; \theta). \quad (2.1)$$

With this setup we can describe most of the possible applications of interest. We now provide three examples to illustrate the prediction function.

³Generally, in particular throughout Section 2.2, we do not distinguish between θ and θ_0 if there is no ambiguity. In Section 2.3 we explicitly use θ_0 to avoid confusion.

Example 2.1. Suppose the time series $\{X_t\}$ follows an AR(1) process given by

$$X_t = \beta X_{t-1} + \varepsilon_t, \quad (2.2)$$

where $|\beta| < 1$ and $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) with $\mathbb{E}[\varepsilon_t] = 0$. The conditional mean of X_{n+1} given X_n is

$$\mu_{n+1} := \mathbb{E}[X_{n+1}|X_n] = \beta X_n. \quad (2.3)$$

Using the prediction function we can then write $\mu_{n+1} = \psi(X_n, X_{n-1}, \dots; \theta) = \beta X_n$ with $\theta = \beta$.

A more precarious example, due to its large popularity, is the conditional variance in a generalized autoregressive conditional heteroskedasticity (GARCH) model (Engle, 1982; Bollerslev, 1986). Whereas in the previous AR(1) case it suffices to condition on the terminal observation, the subsequent Example 2.2 is more extreme as the entire sample contains information about the object of interest.

Example 2.2. Let $\{X_t\}$ follow a GARCH(1, 1) process given by $X_t = \sigma_t \varepsilon_t$ with

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.4)$$

where $\omega > 0$, $\alpha \geq 0$, $1 > \beta \geq 0$ and $\{\varepsilon_t\}$ are i.i.d. with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$. The model's recursive structure implies

$$\sigma_{n+1}^2 = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{\infty} \beta^k X_{n-k}^2. \quad (2.5)$$

It follows directly from (2.5) that

$$\sigma_{n+1}^2 = \psi(X_n, X_{n-1}, \dots; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{\infty} \beta^k X_{n-k}^2,$$

with $\theta = (\omega, \alpha, \beta)'$, and $\Theta \subset (0, \infty) \times [0, \infty) \times [0, 1)$.

The next example shows that for a large class of models the prediction function can be written in the form of Definition 2.1.

Example 2.3. Following Boussama, Fuchs, and Stelzer (2011), consider a Markov

chain of the form

$$S_t = \varphi(S_{t-1}, X_t; \theta), \quad t = 1, 2, \dots \quad (2.6)$$

where φ is some map $\varphi : \mathbb{R}^a \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}^a$. Whereas X_t is observable by the researcher at time t , S_t may be unobservable or only partially observable. The object of interest ψ_{n+1} is typically a function of the state of the Markov chain S_n , such that $\psi_{n+1} = \pi(S_n; \theta)$ for some function π . Through the recursion in (2.6), this is in turn a function of the past of X_n , such that we may write

$$\psi_{n+1} = \pi(S_n; \theta) = \psi_{n+1}(X_n, X_{n-1}, \dots; \theta).$$

Many stochastic processes are in fact Markov processes, including ARMA and GARCH models, several GARCH extensions such as Zakoian's (1994) threshold GARCH, and the set of *observation driven models* considered by Blasques, Koopman, Łasak, and Lucas (2016).

Note that in many cases, such as the GARCH(1,1) of Example 2.2, the prediction function actually depends on the infinite past of the series. In order to express (an approximation of) the prediction function in terms of observable variables only, we would need to replace X_t by s_t for all $t < 1$, where $\{s_t\}$ is a sequence of (arbitrary) constants to which we refer as starting or initial values. For a fixed n , we accordingly define an approximate prediction function $\psi_{n+1}^s : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}$ that is only a function of observable variables and the parameter as

$$\psi_{n+1}^s(\mathbf{X}_{1:n}; \theta) := \psi(X_n, X_{n-1}, \dots, X_1, s_0, s_{-1}, \dots; \theta), \quad (2.7)$$

where $\mathbf{X}_{1:n} = (X_1, \dots, X_n)'$. Note that, given the varying input of the left-hand side in (3.2), we now actually have a sequence of (varying) functions for $n \in \mathbb{N}$.

In many cases the values far in the past are negligible for a wide range of values for $\{s_t\}$. Consequently, ψ_{n+1}^s will be close to ψ_{n+1} . This property can be shown to hold for many different processes including the ones in the examples. We formalize the exact condition we need regarding the negligibility of the starting values in Assumption 2.1.(ii).

Although the prediction function typically represents a conditional object, we have not conditioned on anything yet in the definition. We therefore now extend the analysis by formally conditioning on observing a particular sample. Let $\mathbf{x}_{1:n} =$

$(x_1, \dots, x_n)'$ denote a specific sample path of $\mathbf{X}_{1:n}$. Throughout the chapter, we will discriminate between random variables and their realized counterparts by writing the former in capital and the latter in lowercase letters to avoid ambiguity.

As we will consider sample splitting later on, we define notation that also allows for conditioning on only a subsample. For that purpose, let $t_1 : t_2$ denote the (sub-)period from t_1 up to t_2 , and correspondingly $\mathbf{X}_{t_1:t_2} = (X_{t_1}, \dots, X_{t_2})'$ for any integers $1 \leq t_1 \leq t_2 \leq n$, with a corresponding definition for the observed subsample $\mathbf{x}_{t_1:t_2}$. Furthermore, let $\mathbf{X}_{t_1:n}^c = (c_1, \dots, c_{t_1-1}, X_{t_1}, \dots, X_n)'$ denote the vector where all non-considered subsamples are replaced by a sequence of constants $\{c_t\}$, in a similar way as we did for the starting values. We can now formally define the conditional prediction function.

Definition 2.2. The *prediction function conditional on observing $\mathbf{X}_{t_1:n} = \mathbf{x}_{t_1:n}$* is defined as

$$\psi_{n+1|t_1:n} := \psi_{n+1}^s(\mathbf{x}_{t_1:n}^c; \theta). \quad (2.8)$$

Note that we phrase the conditional prediction function directly in terms of the approximate prediction function ψ_{n+1}^s rather than the true prediction function. We take this “shortcut” because we cannot observe x_0, x_{-1}, \dots , so we cannot condition on those values anyway. Therefore, the “true” conditional object (which we might represent as $\psi_{n+1|-\infty:n}$), is, from an applied point of view, only the theoretical benchmark.

Example 2.1. (*continued*) For the conditional mean of an AR(1) process, conditioning only on the terminal observation $X_n = x_n$ suffices; that is, for any $t_1 \geq 1$ and any sequence $\{c_t\}$, we have that

$$\psi_{n+1|t_1:n} = \psi_{n+1}^s(\mathbf{x}_{t_1:n}^c; \theta) = \beta x_n = \psi_{n+1}^s(\mathbf{x}_{n:n}^c; \theta) = \psi_{n+1|n}. \quad (2.9)$$

Example 2.2. (*continued*) For objects such as the conditional variance for the GARCH(1,1), the conditioning set and the sequence $\{c_t\}$ make a difference, as

$$\begin{aligned} \psi_{n+1|t_1:n} = \psi_{n+1}^s(\mathbf{x}_{t_1:n}^c; \theta) &= \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{n-t_1} \beta^k x_{n-k}^2 \\ &+ \alpha \beta^{n-t_1} \sum_{k=1}^{t_1-1} \beta^k c_{t_1-k}^2 + \alpha \beta^n \sum_{k=0}^{\infty} \beta^k s_{-k}^2, \end{aligned} \quad (2.10)$$

which differs depending on the choice of t_1 . However, as will be shown later, with

an appropriate choice of t_1 , our Assumption 2.1.(ii) on the negligibility of the initial condition, also implies that the difference between $\psi_{n+1|t_1:n}$ and $\psi_{n+1|1:n}$ becomes negligible asymptotically.

Before introducing estimators for θ let us discuss the objects we want to construct inference for. In principle there are two unknown objects one could develop statistical intervals for: $\psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)$ (or slightly more generally $\psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)$) and $\psi_{n+1}^s(\mathbf{x}_{t_1:n}^c; \theta)$. For a GARCH(1,1), for instance, the first would read as

$$\psi_{n+1}^s(\mathbf{X}_{1:n}; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{n-1} \beta^k X_{n-k}^2 + \alpha \beta^n \sum_{k=0}^{\infty} \beta^k s_{-k}^2$$

whereas the second with $t_1 = 1$ reads as

$$\psi_{n+1}^s(\mathbf{x}_{1:n}; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{n-1} \beta^k x_{n-k}^2 + \alpha \beta^n \sum_{k=0}^{\infty} \beta^k s_{-k}^2$$

or more generally, if t_1 is not taken to be equal to one, as in (2.10). While statistical intervals for both objects can be constructed we focus here on conditional inference, i.e. on intervals for $\psi_{n+1|t_1:n} = \psi_{n+1}^s(\mathbf{x}_{t_1:n}^c; \theta)$. In a time series context intervals for $\psi_{n+1|t_1:n}$ are motivated by the *relevance property* of Kabaila (1999) which postulates that intervals should relate to what actually happened during the sample period opposed to what might have happened. Indeed, intervals for $\psi_{n+1|t_1:n}$ can theoretically be shown to be considerably shorter than the intervals for their unconditional counterparts. While the unconditional objects might lead to a conceptually easier analysis, our focus on the conditional objects is not only theoretically but also empirically relevant.

Estimating the Prediction Function

As θ is unobserved, we need to estimate it. We assume that the estimator is based on a subsample $1 : n_E$ (with $1 \leq n_E \leq n$) of the process $\{X_t^E\}$ which is potentially a *different* sample than $\{X_t\}$ that arises in the prediction function. The estimator of θ based on $\mathbf{X}_{1:n_E}^E = (X_1^E, \dots, X_{n_E}^E)'$ will be denoted by $\hat{\theta}(\mathbf{X}_{1:n_E}^E)$. The introduction of $\{X_t^E\}$ serves three purposes: first, using a different process allows us to formulate the two-independent-processes argument where $X_t^E = Y_t$, with $\{Y_t\}$ independent of $\{X_t\}$, $n_E = n$ and an interval is constructed for $\psi_{n+1|1:n}$. Second, it will allow us to discuss the standard approach where $X_t^E = X_t$, $n_E = n$,

and an interval is constructed for $\psi_{n+1|1:n}$. Please note already here that this means that the same variables that arise in the prediction function are also used for estimating θ . Third, it allows us to define the sample splitting approach which we study here. In this approach $X_t^E = X_t$ and an interval is constructed for $\psi_{n+1|n_P:n}$ with $n_E < n_P$ (with $1 < n_P \leq n$) such that in contrast to the standard approach different subsamples are used for prediction and estimation.

Before we illustrate why the standard approach is problematic for constructing and evaluating conditional intervals, we need to define the final building block of prediction function estimation: the conditional prediction function estimator.

Definition 2.3. Let $1 \leq n_P \leq n$. Define the *prediction function estimator conditional on observing $\mathbf{X}_{n_P:n} = \mathbf{x}_{n_P:n}$* as

$$\widehat{\psi}_{n+1|n_P:n} := \widehat{\psi}_{n+1}(\mathbf{x}_{n_P:n}^c, \mathbf{X}_{1:n_E}^E) = \psi_{n+1}^s(\mathbf{x}_{n_P:n}^c, \hat{\theta}(\mathbf{X}_{1:n_E}^E)). \quad (2.11)$$

Note that in the above definition we do not condition on the sample $\mathbf{X}_{1:n_E}^E = \mathbf{x}_{1:n_E}^E$ that is used to estimate θ . The reason for not conditioning on $\mathbf{X}_{1:n_E}^E = \mathbf{x}_{1:n_E}^E$ is that the goal is to preserve the randomness in the second argument of ψ_{n+1}^s , i.e. in $\hat{\theta}(\mathbf{X}_{1:n_E}^E)$, and consequently in $\widehat{\psi}_{n+1|n_P:n}$. Hence, if this goal is achieved we can use the (non-degenerate) conditional (on $\mathbf{X}_{n_P:n} = \mathbf{x}_{n_P:n}$) distribution of $\widehat{\psi}_{n+1|n_P:n}$ to construct confidence intervals for $\psi_{n+1|n_P:n}$. Having this said let us have a closer look at the standard approach. As mentioned above in the standard approach one has $X_t^E = X_t$, $n_P = 1$ and $n_E = n$. Hence, denoting by $\widehat{\psi}_{n+1|1:n}^{STA}$ the “standard” estimator of the prediction function conditional on observing $\mathbf{X}_{1:n} = \mathbf{x}_{1:n}$, it becomes

$$\widehat{\psi}_{n+1|1:n}^{STA} := \widehat{\psi}_{n+1}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) = \psi_{n+1}^s(\mathbf{x}_{1:n}, \hat{\theta}(\mathbf{x}_{1:n})). \quad (2.12)$$

Notice that there is no capital \mathbf{X} in (2.12) because there is only one sample and one typically conditions on all values of this sample. Hence, (2.12) is non-random and thus does not have a distribution that could be used to construct intervals. Instead, to still be able to construct a “standard-looking” interval in practice, researchers typically implicitly rely on the (approximate) quantiles of the estimator

$$\widehat{\psi}_{n+1|1:n}^{STA*} := \widehat{\psi}_{n+1}(\mathbf{x}_{1:n}, \mathbf{X}_{1:n}) = \psi_{n+1}^s(\mathbf{x}_{1:n}, \hat{\theta}(\mathbf{X}_{1:n})). \quad (2.13)$$

It is well understood in the literature that considering the sample as random and non-random at the same time as in (2.13) does not provide a fully satisfactory

justification of the intervals used in practice. For the readers not so familiar with the problem just discussed we provide two examples that both illustrate the problem arising from (2.13). The examples illustrate that the severity of the problem may vary; ranging from only complicating the analysis (Example 2.1) to making the analysis impossible (Example 2.2).

Example 2.1. (*continued*) For the AR(1), we know from (2.9) that $\psi_{n+1|1:n} = \psi_{n+1}^s(\mathbf{x}_{1:n}, \theta) = \beta x_n$. Estimating β by OLS, say $\hat{\beta}(\mathbf{X}_{1:n})$, the estimator in (2.13) becomes

$$\widehat{\psi}_{n+1|1:n}^{STA*} = \psi_{n+1}(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{X}_{1:n})) = \psi_{n+1}^s(\mathbf{x}_{1:n}, \hat{\beta}(\mathbf{X}_{1:n})) = \hat{\beta}(\mathbf{X}_{1:n})x_n. \quad (2.14)$$

Note the discrepancy in treating the terminal observation as random in the estimation sample, yet fixed for the prediction sample. To construct an interval for βx_n , one uses

$$\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2) \quad (2.15)$$

with $\sigma_\beta^2 = 1 - \beta^2$ (cf. Hamilton, 1994, p. 215) and that one can estimate the variance of this normal distribution by $\hat{\sigma}_\beta^2(\mathbf{X}_{1:n}) = 1 - \hat{\beta}(\mathbf{X}_{1:n})^2$. Then an interval for βx_n is typically constructed the following way:

$$\hat{\beta}(\mathbf{X}_{1:n})x_n \pm \Phi^{-1}(\gamma/2) x_n \hat{\sigma}_\beta(\mathbf{X}_{1:n})/\sqrt{n}, \quad (2.16)$$

where Φ^{-1} denotes the standard normal quantile function. However, the interval in (2.16) is hard to interpret as the terminal observation is treated simultaneously as fixed and random. In essence, researchers typically approximate the distribution of $\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n}) - \beta)x_n$ instead of the conditional distribution of $\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n}) - \beta)X_n$ given $X_n = x_n$. The approximation of the latter appears rather cumbersome because even the rather simple condition $X_n = x_n$ has an influence on the whole series $\mathbf{X}_{1:n}$ (Kreiss, 2016). Despite the challenge, Phillips (1979) obtains such approximation based on Edgeworth expansions in the case of $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$.

Example 2.2. (*continued*) For the conditional variance of the GARCH(1,1), the standard estimator of the prediction function conditional on $\mathbf{X}_{1:n} = \mathbf{x}_{1:n}$, is given

by

$$\begin{aligned}\hat{\sigma}_{n+1|1:n}^{2STA} &= \psi_{n+1}^s(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{x}_{1:n})) \\ &= \frac{\hat{\omega}(\mathbf{x}_{1:n})}{1 - \hat{\beta}(\mathbf{x}_{1:n})} + \hat{\alpha}(\mathbf{x}_{1:n}) \sum_{k=0}^{n-1} \hat{\beta}(\mathbf{x}_{1:n})^k x_{n-k}^2 \\ &\quad + \hat{\alpha}(\mathbf{x}_{1:n}) \hat{\beta}(\mathbf{x}_{1:n})^n \sum_{k=0}^{\infty} \hat{\beta}(\mathbf{x}_{1:n})^k s_{-k}^2,\end{aligned}\tag{2.17}$$

where $\hat{\theta}(\mathbf{x}_{1:n}) = (\hat{\omega}(\mathbf{x}_{1:n}), \hat{\alpha}(\mathbf{x}_{1:n}), \hat{\beta}(\mathbf{x}_{1:n}))'$ is some estimate for θ depending on $\mathbf{x}_{1:n}$. Clearly, (2.17) illustrates for the GARCH(1,1) the above mentioned problem that the standard estimator is not random (after conditioning). For the GARCH(1,1) the estimator in (2.13) whose quantiles are used for an interval reads as

$$\begin{aligned}\hat{\sigma}_{n+1}^{2STA*} &= \psi_{n+1}(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{X}_{1:n})) \\ &= \frac{\hat{\omega}(\mathbf{X}_{1:n})}{1 - \hat{\beta}(\mathbf{X}_{1:n})} + \hat{\alpha}(\mathbf{X}_{1:n}) \sum_{k=0}^{n-1} \hat{\beta}(\mathbf{X}_{1:n})^k x_{n-k}^2 \\ &\quad + \hat{\alpha}(\mathbf{X}_{1:n}) \hat{\beta}(\mathbf{X}_{1:n})^n \sum_{k=0}^{\infty} \hat{\beta}(\mathbf{X}_{1:n})^k s_{-k}^2.\end{aligned}\tag{2.18}$$

This quantity exemplifies for the GARCH(1,1) that the complete sample is regarded as random and non-random at the same time. While for the AR(1) this complicated the analysis, yet not made it impossible, the dependence on the complete sample here makes it difficult to use this quantity to make meaningful probabilistic statements.

2.2.2 Argument of Two Independent Processes

The argument of two independent processes can at least be traced back to Akaike (1969), who studies the prediction of AR time series. It reoccurs in Lewis and Reinsel (1985, p. 394): “...the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure.” The same argument also appears in Lütkepohl (2005, p. 95) and in Dufour and Taamouti (2010). Let $\{Y_t\}$ be a process independent of $\{X_t\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\{Y_t\}$ having the same stochastic structure as $\{X_t\}$. In addition to the sample $\mathbf{X}_{1:n}$ of

the process $\{X_t\}$, suppose there is a sample $\mathbf{Y}_{1:n} := (Y_1, \dots, Y_n)'$ of the process $\{Y_t\}$ that we use as estimation sample, that is $\mathbf{X}_{1:n}^E = \mathbf{Y}_{1:n}$. In this situation we denote the conditional prediction function estimator of Definition 2.3 by $\widehat{\psi}_{n+1|1:n}^{2IP}$ and it equals

$$\widehat{\psi}_{n+1|1:n}^{2IP} := \widehat{\psi}_{n+1}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) = \psi_{n+1}^s(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{Y}_{1:n})). \quad (2.19)$$

Notice that (2.19) does not have the same shortcoming as (2.13) because even if we consider $\mathbf{x}_{1:n}$ to be known we can nevertheless consider $\hat{\theta}(\mathbf{Y}_{1:n})$ to be random and can hence use its distribution to construct intervals. Throughout this chapter, we call (2.19), the 2IP (two independent processes) estimator. Then, a conditional interval $I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n})$ can be based on the (approximate) quantiles of $\widehat{\psi}_{n+1|1:n}^{2IP}$. It satisfies

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) \ni \psi_{n+1|1:n} \mid \mathbf{X}_{1:n} = \mathbf{x}_{1:n}\right] \stackrel{(\approx)}{=} 1 - \gamma, \quad (2.20)$$

with the approximate sign indicating asymptotic equivalence. Note that the independence implies that the distribution of $\mathbf{Y}_{1:n}$ in (2.20) does not depend on the realization $\mathbf{x}_{1:n}$, yet the statement does depend on $\mathbf{x}_{1:n}$ because the interval depends on it (for the AR(1) this can be directly seen from (2.16) when replacing $\mathbf{X}_{1:n}$ by $\mathbf{Y}_{1:n}$). Although the 2IP approach is statistically sound, it assumes two independent processes with the same stochastic structure. Phillips (1979) points out that the assumption “is quite unrealistic in practical situations” (p. 241). Indeed, it is difficult to imagine this assumption to be satisfied in any real-life application beyond experimental settings. Moreover, as only one sample realization is available, to compute the estimate of the interval $I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n})$ it is frequently suggested to take $\mathbf{Y}_{1:n} = \mathbf{x}_{1:n}$, violating the independence assumption. Thus, the 2IP approach appears to be a rather questionable justification for the usual interval, and as such, in this work we provide an alternative, realistic, justification of (asymptotically equivalent) intervals based on sample splitting.

2.2.3 Sample-split Estimation

An intuitive motivation for the sample-split approach is the successive decline of the influence of past observations present in a substantial class of time series models. This property permits to split our sample into two (asymptotically) independent subsamples. Consider the end point of the estimation sample, n_E , and the

starting point of the prediction sample, n_P satisfying $1 < n_E < n_P \leq n$, such that the two samples are non-overlapping. In this situation we denote the conditional prediction function estimator of Definition 2.3 by $\widehat{\psi}_{n+1|n_P:n}^{SPL}$ and it is given by

$$\widehat{\psi}_{n+1|n_P:n}^{SPL} := \widehat{\psi}_{n+1}(\mathbf{x}_{n_P:n}^c, \mathbf{X}_{1:n_E}) = \psi_{n+1}^s(\mathbf{x}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E})). \quad (2.21)$$

Throughout the chapter, we call (2.21) the SPL estimator (due to splitting). Similar to the two sample approach, we can consider the first argument of ψ_{n+1}^s in (2.21) as given and the second argument as random since the subsamples are non-overlapping. A conditional interval $I_\gamma^{SPL}(\mathbf{x}_{n_P:n}, \mathbf{X}_{1:n_E})$ can be constructed such that

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{x}_{n_P:n}, \mathbf{X}_{1:n_E}) \ni \psi_{n+1|n_P:n} \mid \mathbf{X}_{n_P:n} = \mathbf{x}_{n_P:n}\right] \stackrel{(\approx)}{=} 1 - \gamma. \quad (2.22)$$

This statement does make sense as there is still randomness in $\hat{\theta}(\mathbf{X}_{1:n_E})$ since $\mathbf{X}_{1:n_E}$ is not conditioned upon, yet the last $n - n_P + 1$ values of $\{X_t\}_{t=1}^n$ are fixed such that their randomness is not taken into account. Similar to (2.20), the statement in (2.22) does depend on $\mathbf{x}_{n_P:n}$ and in contrast to $\mathbf{x}_{1:n}$ in (2.20) the realization $\mathbf{x}_{n_P:n}$ may influence the distribution of $\mathbf{X}_{1:n_E}$. However, as said at the beginning of this subsection the idea of the sample-split approach is that this dependence will vanish asymptotically.

Remark 2.1. In Section 2.3 we will discuss how n_E and n_P should be chosen from an asymptotic perspective to ensure that our regularity conditions are fulfilled. As we only consider sample splitting as a theoretical approach to validate commonly constructed conditional confidence intervals, these asymptotic guidelines are sufficient for our purposes and we do not have to consider how to choose n_E and n_P in practice. Of course, one could use the sample-split approach in practice to construct confidence intervals. While one would gain (near) independence between the two subsamples, this would come at a cost of less estimation precision as fewer observations are used for parameter estimation. For the Gaussian AR(1) setting, Phillips (1979) derives asymptotic expansions for the case where, in our notation, $n_E = n - l$ and $n_P = n$ for some $l \geq 0$, showing that even in this simple case there is indeed a trade-off as described above and the optimal choice of l is unclear. An interesting extension of our analysis would therefore be to investigate the optimal choices of n_E and n_P to achieve the most accurate confidence intervals in small samples. However, this choice is likely to be highly dependent on the specific

model and as such would have to be investigated on a case-by-case basis. This is therefore outside the scope of the current work.

2.3 Asymptotic Justification

In this section, we connect the sample-split procedure of Section 2.2.3 with the two-independent-samples approach of Section 2.2.2. First, in Section 2.3.1, we show that the notion of weak convergence is inadequate to study asymptotic closeness for objects that vary over time and discuss the concept of merging. Then, in Section 2.3.2 we link the 2IP and the SPL estimator by proving that their conditional distributions merge in probability (Theorem 2.1). Thereafter, in Section 2.3.3, we construct asymptotically valid intervals (Theorem 2.2) and show that the sample-split intervals coincide asymptotically with the intervals commonly constructed by practitioners (Theorem 2.3). Last, in Section 2.3.4, we state intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

2.3.1 Merging

To illustrate the inappropriateness of weak convergence in the context considered here, we revisit Example 2.1 for the 2IP approach and the SPL approach, which shows that studying asymptotic closeness between conditional distributions is often complicated by the absence of a limiting distribution.

Example 2.1. (*continued*) For the 2IP approach (2.15) implies that $\sqrt{n}(\hat{\beta}(\mathbf{Y}_{1:n}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2)$ and it entails that $\sqrt{n}(\hat{\beta}(\mathbf{Y}_{1:n}) - \beta)x$ converges weakly to $N(0, \sigma_\beta^2 x^2)$ for any fixed $x \neq 0$. Further, the result suggests that the conditional distribution of $\sqrt{n}(\hat{\beta}(\mathbf{Y}_{1:n}) - \beta)X_n$ given $X_n = x_n$, which is just the distribution of $\sqrt{n}(\hat{\beta}(\mathbf{Y}_{1:n}) - \beta)x_n$, is asymptotically close to $N(0, \sigma_\beta^2 x_n^2)$. Similarly, for the SPL approach with $n_E/n \rightarrow 1$ we have $\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n_E}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2)$ which suggests as well (if the gap between n_E and n is large enough which will be formally specified below) that the conditional distribution of $\sqrt{n_E}(\hat{\beta}(\mathbf{X}_{1:n_E}) - \beta)X_n$ given $X_n = x_n$ is also close to $N(0, \sigma_\beta^2 x_n^2)$. For both approaches the approximating distribution $N(0, \sigma_\beta^2 x_n^2)$ varies with n through the terminal realization x_n . Note that the concept of weak convergence is not applicable in this context to characterize this

asymptotic closeness, as it requires a (fixed) limiting distribution, which is absent here.

Next, we discuss what closeness means in the absence of a limiting distribution. To do so, first recall that weak convergence of a sequence of cdfs $\{F_n\}$ on \mathbb{R}^k with $k \in \mathbb{N}$, i.e. $F_n(\tau) \rightarrow F(\tau)$ for all continuity points τ of F , can alternatively be defined by $d_{BL}(F_n, F) \rightarrow 0$. Here d_{BL} denotes the bounded Lipschitz metric defined by

$$d_{BL}(F, G) = \sup \left\{ \left| \int f d(F - G) \right| : \|f\|_{BL} \leq 1 \right\}, \quad (2.23)$$

where for any real-valued function f on \mathbb{R}^k one puts $\|f\|_{BL} = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$, with $\|\cdot\|$ denoting the Euclidean norm, i.e. $\|A\| = \sqrt{\text{tr}(A'A)}$ for any vector or matrix A . Following Dudley (2002) (see D'Aristotile, Diaconis, and Freedman (1988) and Davydov and Rotar (2009) for related work) we state the following definition.

Definition 2.4. (*Merging*) Two sequences of cdfs $\{F_n\}$ and $\{G_n\}$ are said to *merge* if and only if $d_{BL}(F_n, G_n) \rightarrow 0$ as $n \rightarrow \infty$.

Note that weak convergence can be seen as a special case of merging with $G_n = G$ for all $n \in \mathbb{N}$. While merging is appropriate to capture the asymptotic closeness of the conditional distribution of $\sqrt{n}(\hat{\beta}(\mathbf{Y}_{1:n}) - \beta)X_n$ and $N(0, \sigma_\beta^2 x_n^2)$ for a given sample $X_n = x_n$, we now extend the concept in a way that allows us to deal with asymptotic closeness when we do not condition on a particular sample. The necessity of this definition can again be exemplified by the AR(1), which also illustrates how we will deal with the dependence of the statements in (2.20) and in (2.22) on the sample that we mentioned below these equations. For instance, in Example 2.1 as described at the beginning of this section, the goal would be to formalize a statement like ‘when n is large, the probability of all x_n such that the distribution of $\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n}) - \beta)x_n$ merges with that of $\sqrt{n_E}(\hat{\beta}(\mathbf{X}_{1:n_E}) - \beta)x_n$ is approximately equal to one’. We now first introduce the conditional distributions of the 2IP and the SPL estimator in the general case and then give the definition capturing what we just illustrated for the AR(1).

Let m_n be a sequence of normalizing constants with $m_n \rightarrow \infty$ (e.g. $m_n = \sqrt{n}$). For any $t_1 \geq 1$, we define the sub σ -algebra $\mathcal{I}_{t_1:n} = \sigma(X_t : t_1 \leq t \leq n)$. We denote

the *conditional cdfs* of the 2IP and SPL estimator by

$$F_n^{2IP}(\tau | \mathcal{I}_{1:n}) := \mathbb{P}[m_n(\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{Y}_{1:n}) - \psi_{n+1}) \leq \tau | \mathcal{I}_{1:n}] \quad (2.24)$$

$$F_n^{SPL}(\tau | \mathcal{I}_{n_P:n}) := \mathbb{P}[m_n(\hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}, \mathbf{X}_{1:n_E}) - \psi_{n+1}) \leq \tau | \mathcal{I}_{n_P:n}], \quad (2.25)$$

respectively, so that by specifying an event of $\mathcal{I}_{1:n}$ and $\mathcal{I}_{n_P:n}$, we see that (2.24) and (2.25) are just the centered and scaled distributions of (2.19) and (2.21), respectively. Please note that (2.25) actually also depends on c , see (2.21), but since our assumptions will ensure that this dependence vanishes asymptotically we prefer to suppress the dependence on c here.

Remark 2.2. Although not explicitly mentioned above we consider (2.24) and (2.25) to be regular conditional cdfs, i.e. we assume that $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})(\omega)$ and $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})(\omega)$ are cdfs for every $\omega \in \Omega$; for the exact definition and the existence see Dudley (2002, Section 10.2).

We can now define merging in probability (we do so without explicitly using the conditional cdfs of the 2IP and the SPL approach).

Definition 2.5. (*Merging in probability*) Two sequences of conditional cdfs $\{F_n\}$ and $\{G_n\}$ are said to *merge in probability* if and only if $d_{BL}(F_n, G_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$, where “ \xrightarrow{P} ” denotes “convergence in probability”.

2.3.2 Merging of 2IP and SPL in Probability

Here, we give conditions such that the conditional cdfs of the 2IP and SPL estimator merge. Clearly, the conditional confidence intervals are functions of these distributions so that their merging is a building block for the study of the conditional confidence intervals based on them. The conditions we give are divided into three parts. Roughly speaking, the first part (general assumptions) makes sure that the function we want to predict is well behaved and that we can estimate the parameter it depends on. The second part (two independent processes) and third part (SPL estimator) guarantee that these assumptions are met by the 2IP and the SPL method. To write the conditions in compact form we employ the usual stochastic order symbols O_p and o_p . We assume that θ_0 belongs to the interior of Θ , i.e. $\theta_0 \in \overset{\circ}{\Theta}$, and we denote the set of all bounded, real-valued Lipschitz functions on \mathbb{R}^r by $BL = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} < \infty\}$. We start with the general assumptions.

Assumption 2.1. (*General assumptions*)

- (i) (*Estimator*) $m_n(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0) \xrightarrow{d} G_\infty$ as $n \rightarrow \infty$ for some cdf $G_\infty : \mathbb{R}^r \rightarrow [0, 1]$;
- (ii) (*Differentiability*) $\psi(\cdot; \theta)$ is continuous on Θ and twice differentiable on $\mathring{\Theta}$;
- (iii) (*Gradient*) $\left\| \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| = O_p(1)$;
- (iv) (*Hessian*) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ for some open neighborhood $\mathcal{V}(\theta_0)$ around θ_0 ;
- (v) (*Initial condition*) Given sequences $\{s_t\}$ and $\{c_t\}$, we have

$$\begin{aligned} m_n(\psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) - \psi(X_n, X_{n-1}, \dots; \theta_0)) &= o_p(1), \\ \left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| &= o_p(1), \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| &= o_p(1) \end{aligned}$$

for any $t_1 \geq 1$ such that $(n - t_1)/l_n \rightarrow \infty$ as $n \rightarrow \infty$ and for some model-specific l_n with $l_n \rightarrow \infty$.

Assumption 2.1.(i) implies the existence of a limiting distribution for the parameter estimator. The differentiability assumption in 2.1.(ii) plus the boundedness Assumptions 2.1.(iii) ensure that the scaled prediction function estimators can accurately be approximated by a Taylor expansion; see Lemma 2.1 for details. Assumption 2.1.(v) with $t_1 = 1$ ensures the negligibility of the starting values when using the full-sample for prediction, while taking $t_1 = n_P$ ensures that this extends to the case where additionally X_1, \dots, X_{n_P-1} are replaced by constants, i.e. where only the subsample (X_{n_P}, \dots, X_n) is used for prediction. This assumption implicitly limits the choice of n_P ; as replacing past values of X_t for $t < n_P$ by arbitrary constants should have a negligible effect, $n - n_P$ needs to increase faster than some lower bound l_n . For models exhibiting an exponential decay in memory, it typically suffices to take $l_n = \log n$ (see Chapter 3, Equation (3.13)). For the 2IP estimator, we additionally need the two-independent-processes assumption, which is formalized in Assumption 2.2.

Assumption 2.2. (*Two independent processes*)

- (i) (*Existence*) $\{Y_t\}$ is a process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, distributed as $\{X_t\}$;
- (ii) (*Independence*) $\{Y_t\}$ is independent of $\{X_t\}$.

For the SPL estimator we replace the two-independent-processes assumption by a stationarity and a weak dependence condition, which allows to split our sample into two (asymptotically) independent and identical subsamples. In addition we need an assumption on n_P and n_E as functions of n , that is $n_E(n)$ and $n_P(n)$.

Assumption 2.3. (*SPL estimator*)

- (i) (*Rates*) The functions $n_P : \mathbb{N} \rightarrow \mathbb{N}$ and $n_E : \mathbb{N} \rightarrow \mathbb{N}$ satisfy $n_E(n) < n_P(n)$ for all n , while $\frac{n - n_P(n)}{l_n} \rightarrow \infty$ and $m_{n_E(n)}/m_n \rightarrow 1$ as $n \rightarrow \infty$;
- (ii) (*Strict stationarity*) $\{X_t\}$ is a strictly stationary process;
- (iii) (*Weak dependence*) $\{X_t\}$ satisfies for each $h \in BL$

$$\int h d\left(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_{n_E}^{SPL}\right) \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty,$$

where $G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ denotes the conditional cdf of $m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$ given $\mathcal{I}_{n_P:n}$ and $G_{n_E}^{SPL}$ the corresponding unconditional cdf.

The subsample size assumption in 2.3.(i) ensures that the number of observations used for conditioning is increasing, which along with the negligibility of the initial conditions implies that the truncation of the prediction function is negligible. Furthermore, the sample size used for estimation should increase fast enough that the respective scaling of the 2IP and SPL estimators, m_n and m_{n_E} respectively, are asymptotically identical. If m_n increases no faster than a polynomial rate, which is generally the case, it is sufficient that $n_E/n \rightarrow 1$ for $m_{n_E}/m_n \rightarrow 1$ to hold.

The stationarity assumption in 2.3.(ii) can actually be relaxed; what matters is that the conditions in Assumption 2.1 are still true if only a subsample is considered. In particular, we need that $m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0) \xrightarrow{d} G_\infty$, which - along with the assumptions on gradient and Hessian - is certainly satisfied under stationarity. However, in general the assumption will be far too strict; here we use it simply to have a clear, interpretable assumption rather than a list of high-level assumptions that are difficult to interpret. The weak dependence condition in 2.3.(iii) is met

by numerous Markov processes. Intuitively, (X_1, \dots, X_{n_E}) and (X_{n_P}, \dots, X_n) approach independence as their temporal distance $n_P - n_E$ increases. We illustrate a particular case in the Remark 2.3.

Remark 2.3. Suppose $\{X_t\}$ is strong mixing (cf. Doukhan, 1995) and let α denote the strong mixing coefficient. For $h \in BL$ and for all $\epsilon > 0$, Markov's and Ibragimov's inequality (cf. Hall and Heyde, 1980, Theorem A.5) imply

$$\begin{aligned} & \mathbb{P} \left[\left| \int h d \left(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_{n_E}^{SPL} \right) \right| \geq \epsilon \right] \\ & \leq \frac{1}{\epsilon} \mathbb{E} \left| \int h d \left(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_{n_E}^{SPL} \right) \right| \\ & = \frac{1}{\epsilon} \text{Cov} \left[h \left(m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0) \right), \text{sign} \left\{ \int h d \left(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_{n_E}^{SPL} \right) \right\} \right] \\ & \leq \frac{4 \|h\|_{BL}}{\epsilon} \alpha(n_P - n_E). \end{aligned}$$

Taking $n_P - n_E \rightarrow \infty$ such that $\alpha(n_P - n_E) \rightarrow 0$ verifies Assumption 2.3.(iii).

Assumptions 2.1 to 2.3 are met by the AR and GARCH processes considered in Examples 2.1 and 2.2 (with Assumption 2.1.(v) holding for bounded sequences). A detailed verification of each assumption under mild conditions is provided in Chapter 3. We state the following theorem.

Theorem 2.1. (*Merging of 2IP and SPL*) *If Assumptions 2.1 to 2.3 hold, then $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ and $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ merge in probability.*

Having established asymptotic closeness between the conditional cdfs $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ and $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$, we now turn to the construction of asymptotic intervals.

2.3.3 Interval Construction

Henceforth, for any cdf F we write F^{-1} to denote its generalized inverse given by $F^{-1}(u) = \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u \}$. A confidence interval for ψ_{n+1} based on quantiles of (2.24) or (2.25) is typically infeasible as these cumulative distribution functions are unknown for finite n . Here, they are infeasible because roughly they are the distribution functions of some weights which induce merging multiplied by $m_n(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0)$ and $m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$, respectively, where, in general, the distributions of $m_n(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0)$ and $m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$ are unknown in finite samples. Since these are the only unknown distributions an asymptotic

approximation can be based on G_∞ with merging induced by the non-convergent weights. In general, we also need to estimate G_∞ ; see Examples 2.4 and 2.5 below for common approaches. We denote estimators of (2.24) and (2.25) resulting from this approximation by $\widehat{F}_n^{2IP}(\cdot)$ and $\widehat{F}_n^{SPL}(\cdot)$, respectively. In the next subsection we provide explicit expressions when G_∞ is multivariate normal. For the general construction, we refer to relations (2.42) and (2.43) in Appendix 2.A and the explanations preceding these relations. Based on the 2IP approach, we consider an interval of the form

$$I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) = \left[\widehat{\psi}_{n+1|1:n}^{2IP} - \frac{\widehat{F}_n^{2IP}{}^{-1}(1 - \gamma_2)}{m_n}, \widehat{\psi}_{n+1|1:n}^{2IP} - \frac{\widehat{F}_n^{2IP}{}^{-1}(\gamma_1)}{m_n} \right], \quad (2.26)$$

where $\gamma_1, \gamma_2 \in [0, 1)$ satisfy $\gamma = \gamma_1 + \gamma_2$. We typically take $\gamma_1 = \gamma_2 = \gamma/2$ such that the interval is equal-tailed. Similarly, we construct the following sample-split interval:

$$I_\gamma^{SPL}(\mathbf{x}_{nP:n}^c, \mathbf{X}_{1:nE}) = \left[\widehat{\psi}_{n+1|nP:n}^{SPL} - \frac{\widehat{F}_n^{SPL}{}^{-1}(1 - \gamma_2)}{m_n}, \widehat{\psi}_{n+1|nP:n}^{SPL} - \frac{\widehat{F}_n^{SPL}{}^{-1}(\gamma_1)}{m_n} \right]. \quad (2.27)$$

To achieve correct coverage, we need that $\widehat{F}_n^{2IP}(\cdot)$ and $F_n^{2IP}(\cdot|\mathcal{I}_{1:n})$ merge in probability and likewise for SPL. A sufficient condition for this in our setting is that we can consistently estimate the asymptotic distribution of the parameter estimator, G_∞ , by an appropriate estimator. This is formulated in Assumption 2.4 below.

Assumption 2.4. (*CDF estimator*) Let $\widehat{G}_n(\cdot)$ denote a random (r -dimensional) cdf as a function of $\mathbf{X}_{1:n}$, used to estimate G_∞ : i.e. $\int h d\widehat{G}_n(\cdot) \xrightarrow{P} \int h dG_\infty$ as $n \rightarrow \infty$ for all $h \in BL$.

Although we did not explicitly specify in Assumption 2.4 the dependence of \widehat{G}_n on $\mathbf{X}_{1:n}$, it should be understood to hold for any subsample of $\mathbf{X}_{1:n}$ whose size goes to infinity. The verification of Assumption 2.4 is a standard step in asymptotic analysis. The two examples below provide common methods for verifying Assumption 2.4.

Example 2.4. Suppose that G_∞ belongs to some parametric family $\{G_{\theta,\xi}|\theta \in \Theta, \xi \in \Xi\}$, where ξ corresponds to additional nuisance parameters. Then, given

some consistent estimators $\hat{\theta}(\mathbf{X}_{1:n})$ and $\hat{\xi}(\mathbf{X}_{1:n})$ for θ_0 and ξ_0 respectively, it follows from the continuous mapping theorem that $\widehat{G}_n = G_{\hat{\theta}(\mathbf{X}_{1:n}), \hat{\xi}(\mathbf{X}_{1:n})}$ satisfies Assumption 2.4 if $G_{\theta, \xi}$ is continuous in θ and ξ .

Example 2.5. If \hat{G}_n is based on a consistent bootstrap procedure for G_∞ then Assumption 2.4 clearly holds.

The following theorem states the intervals' asymptotic validity.

Theorem 2.2. (*Asymptotic coverage*)

1. (i) Under Assumption 2.1, 2.2 and 2.4, $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ and $\widehat{F}_n^{2IP}(\cdot)$ merge in probability.
- (ii) If in addition $\widehat{F}_n^{2IP}(\cdot)$ is stochastically uniformly equicontinuous, then

$$\mathbb{P} \left[I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) \ni \psi_{n+1} \middle| \mathcal{I}_{1:n} \right] \xrightarrow{P} 1 - \gamma. \quad (2.28)$$

2. (i) Under Assumption 2.1, 2.3 and 2.4, $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ and $\widehat{F}_n^{SPL}(\cdot)$ merge in probability.
- (ii) If in addition $\widehat{F}_n^{SPL}(\cdot)$ is stochastically uniformly equicontinuous, then

$$\mathbb{P} \left[I_\gamma^{SPL}(\mathbf{x}_{n_P:n}^c, \mathbf{X}_{1:n_E}) \ni \psi_{n+1} \middle| \mathcal{I}_{n_P:n} \right] \xrightarrow{P} 1 - \gamma. \quad (2.29)$$

However, the standard approach, motivated by I_γ^{2IP} as in (2.28), computes an interval of the form $I_\gamma^{STA}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) = I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n})$ as only one sample realization is available. This, of course, strongly violates the independence assumption of $\{X_t\}$ and $\{Y_t\}$. Specifically, replacing $\mathbf{Y}_{1:n}$ by $\mathbf{X}_{1:n}$ in equation (2.26), leads to

$$\begin{aligned} & I_\gamma^{STA*}(\mathbf{x}_{1:n}, \mathbf{X}_{1:n}) \\ &= \left[\widehat{\psi}_{n+1|1:n}^{STA*} - \frac{\widehat{F}_n^{STA*}{}^{-1}(1 - \gamma_2)}{m_n}, \widehat{\psi}_{n+1|1:n}^{STA*} - \frac{\widehat{F}_n^{STA*}{}^{-1}(\gamma_1)}{m_n} \right], \end{aligned} \quad (2.30)$$

where $\widehat{F}_n^{STA*}(\cdot)$ is defined in relation (2.44) and the text preceding it. Whereas it is difficult to justify a conditional confidence interval like $I_\gamma^{STA}(\mathbf{x}_{1:n}, \mathbf{X}_{1:n})$ directly due to the lack of randomness, we can provide a justification by characterizing how closely the interval resembles the SPL interval. We establish the asymptotic equivalence, defined in terms of location and (scaled) length, between the two

intervals in the following theorem. Note that, as our characterization of equivalence is probabilistic, we need to introduce the “doubly random” versions of the STA and SPL estimators, where the sample we condition on is considered random. These estimators are denoted as $\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n})$ and $\hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E})$ respectively.

Theorem 2.3. (*Asymptotic equivalence of confidence intervals*)

1. (*Location*) If Assumptions 2.1 to 2.3 hold, then

$$\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) - \hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E}) \xrightarrow{P} 0.$$

2. (*Length*) Under the assumptions of Theorem 2.1 and 2.2 and $\widehat{F}_n^{SPL^{-1}}(\cdot)$ being stochastically pointwise continuous at $u = \gamma_1, 1 - \gamma_2$, we have

$$\widehat{F}_n^{STA^{-1}}(u) - \widehat{F}_n^{SPL^{-1}}(u) \xrightarrow{P} 0. \quad (2.31)$$

The first implication states that the locations of the two intervals coincide asymptotically. The second statement establishes asymptotic closeness of the selected quantiles such that the scaled lengths of the intervals in (2.27) and (2.30) coincide asymptotically. As such, our sample-split interval coincides asymptotically with the standard interval, meaning that the standard interval can be substituted for an (asymptotically) equivalent interval which has a formal justification in terms of conditional coverage. As such, this provides a justification for the intervals commonly constructed in practice without having to rely on the two-independent-processes assumption.

2.3.4 Interval Construction Under Normality

In this subsection we present intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

Assumption 2.5. (*Normality*) Let G_∞ be the cdf of the $N(0, \Upsilon_0)$ distribution with $\Upsilon_0 = \Upsilon(\theta_0, \xi_0)$ and assume there exist $\hat{\Upsilon}(\mathbf{X}_{1:n})$ converging in probability to Υ_0 .

Usually, the covariance estimator is obtained by inserting consistent estimators for θ_0 and ξ_0 into Υ_0 . Following the plug-in principle, we estimate $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ by a

normal distribution with mean 0 and variance

$$\hat{v}_n^{2IP} = \frac{\partial \psi_{n+1}(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{Y}_{1:n}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{Y}_{1:n}) \frac{\partial \psi_{n+1}(x_{1:n}; \hat{\theta}(\mathbf{Y}_{1:n}))}{\partial \theta}$$

such that $\widehat{F}_n^{2IP}(\cdot) = \Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$. Then, the interval in (2.26) simplifies to

$$\begin{aligned} & I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) \\ &= \left[\hat{\psi}_{n+1|1:n}^{2IP} - \frac{\sqrt{\hat{v}_n^{2IP}} \Phi^{-1}(1 - \gamma_2)}{m_n}, \hat{\psi}_{n+1|1:n}^{2IP} - \frac{\sqrt{\hat{v}_n^{2IP}} \Phi^{-1}(\gamma_1)}{m_n} \right]. \end{aligned} \quad (2.32)$$

Similarly, for the sample-split approach we get $\widehat{F}_n^{SPL}(\cdot) = \Phi(\cdot / \sqrt{\hat{v}_n^{SPL}})$ with

$$\hat{v}_n^{SPL} = \frac{\partial \psi_{n+1}(\mathbf{x}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_{1:n_E}) \frac{\partial \psi_{n+1}(\mathbf{x}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E}))}{\partial \theta}$$

such that (2.27) reduces to

$$\begin{aligned} & I_\gamma^{SPL}(\mathbf{x}_{n_P:n}, \mathbf{X}_{1:n_E}) \\ &= \left[\hat{\psi}_{n+1|n_P:n}^{SPL} - \frac{\sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(1 - \gamma_2)}{m_n}, \hat{\psi}_{n+1|n_P:n}^{SPL} - \frac{\sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(\gamma_1)}{m_n} \right]. \end{aligned} \quad (2.33)$$

In Appendix 2.B we show that if the variance estimator is bounded away from zero in probability, e.g. $1/\hat{v}_n^{2IP} = O_p(1)$, then $\widehat{F}_n^{2IP}(\cdot)$ is stochastically uniform equicontinuous. Therefore, the asymptotic validity of both intervals can be deduced from Theorem 2.2.

Corollary 2.1. (*Asymptotic coverage under normality*)

1. (i) Under Assumption 2.1, 2.2 and 2.5, $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ and $\Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$ merge in probability.
- (ii) If in addition $1/\hat{v}_n^{2IP} = O_p(1)$, then

$$\mathbb{P} \left[I_\gamma^{2IP}(\mathbf{x}_{1:n}, \mathbf{Y}_{1:n}) \ni \psi_{n+1} \middle| \mathcal{I}_{1:n} \right] \xrightarrow{p} 1 - \gamma.$$

2. (i) Under Assumption 2.1, 2.3 and 2.5, $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ and $\Phi(\cdot / \sqrt{\hat{v}_n^{SPL}})$ merge in probability.

(ii) If in addition $1/\hat{v}_n^{SPL} = O_p(1)$, then

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{x}_{n_P:n}^c, \mathbf{X}_{1:n_E}) \ni \psi_{n+1} \middle| \mathcal{I}_{n_P:n}\right] \xrightarrow{p} 1 - \gamma.$$

Bounding the variance estimator away from zero in probability to establish that the conditional coverage probability converges to $1 - \gamma$ in probability has intuitive appeal: as \hat{v}_n^{2IP} approaches zero, $N(0, \hat{v}_n^{2IP})$ becomes degenerate while the interval in (2.32) collapses (similar for SPL).

For the standard interval, replacing $\mathbf{Y}_{1:n}$ by $\mathbf{X}_{1:n}$ in (2.32) leads to

$$\begin{aligned} & I_\gamma^{STA*}(\mathbf{x}_{1:n}, \mathbf{X}_{1:n}) \\ &= \left[\hat{\psi}_{n+1|n_P:n}^{STA*} - \frac{\sqrt{\hat{v}_n^{STA} \Phi^{-1}(1 - \gamma_2)}}{m_n}, \hat{\psi}_{n+1|n_P:n}^{STA*} - \frac{\sqrt{\hat{v}_n^{STA} \Phi^{-1}(\gamma_1)}}{m_n} \right] \end{aligned} \quad (2.34)$$

with $\hat{v}_n^{STA} = \frac{\partial \psi_{n+1}(x_{1:n}; \hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_n) \frac{\partial \psi_{n+1}(\mathbf{x}_{1:n}; \hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta}$. In Appendix 2.B we prove that \hat{v}_n^{SPL} is bounded in probability, which in turn implies that the quantile function $\widehat{F}_n^{SPL^{-1}}(\cdot) = \sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(\cdot)$ is stochastically pointwise equicontinuous at any $u \in \mathbb{R}$. Hence, Theorem 2.3 applies. Whereas the first statement of the theorem remains unaffected, its second statement reads as follows under normality.

Corollary 2.2. (*Length under normality*) *Under the assumptions of Theorem 2.1 and Corollary 2.1, we have*

$$\sqrt{\hat{v}_n^{STA} \Phi^{-1}(u)} - \sqrt{\hat{v}_n^{SPL} \Phi^{-1}(u)} \xrightarrow{p} 0$$

for $u = \gamma_1, 1 - \gamma_2$.

2.4 Prediction Intervals

The preceding sections have focused purely on the construction of conditional confidence intervals to account for parameter uncertainty. Regarding prediction, a second source of uncertainty arises, that corresponds to the model's innovation process. In this setting, parameter estimation is typically disregarded in textbooks as the stochastic fluctuation stemming from the estimation procedure is generally dominated by the stochastic fluctuation of the innovations. Although the resulting prediction intervals may be asymptotically valid, they are typically characterized

by under-coverage in finite samples. In response, Pan and Politis (2016a) introduce the concept of *asymptotic pertinence* to evaluate distribution approximations that account for the two sources of randomness, innovation and parameter estimation uncertainty, according to their general orders of magnitude. Whereas Kunitomo and Yamamoto (1985) and Samaranayake and Hasza (1988) study properties of the *unconditional* distribution of the forecast error, we focus on its *conditional* distribution to conform with the relevance property of Kabaila (1999). The fundamental issue also arises when considering prediction if one attempts to account for parameter uncertainty. To illustrate this point, we revisit the introductory examples and write $*$ to denote the convolution operator.⁴

Example 2.1. (*continued*) Prediction intervals for the AR are often constructed around the point estimate for the conditional mean. The conditional distribution of the forecast error decomposes into

$$\begin{aligned} & \mathbb{P}[X_{n+1} - \hat{\beta}(\mathbf{X}_{1:n})X_n \leq \cdot | X_n = x_n] \\ &= \mathbb{P}[\beta X_n - \hat{\beta}(\mathbf{X}_{1:n})X_n \leq \cdot | X_n = x_n] * \mathbb{P}[\varepsilon_{n+1} \leq \cdot], \end{aligned} \quad (2.35)$$

corresponding to estimation and innovation uncertainty, respectively. As argued above, an approximation of $\mathbb{P}[\beta X_n - \hat{\beta}(\mathbf{X}_{1:n})X_n \leq \cdot | X_n = x_n]$ appears rather cumbersome. In the special case of $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$, Phillips (1979, Theorem 3) derives an approximation for (2.35) based on Edgeworth expansions.

Example 2.2. (*continued*) Suppose we are interested in providing a prediction interval for X_{n+1}^2 in the GARCH(1,1). Conditioning on $\mathbf{X}_{1:n} = \mathbf{x}_{1:n}$, a common estimate of X_{n+1}^2 is $\hat{\sigma}_{n+1|1:n}^{2STA}$ as defined in (2.17), since $\sigma_{n+1|1:n}^2$ is its expected value given information up to time n . As

$$\begin{aligned} & \mathbb{P}[X_{n+1}^2 - \hat{\sigma}_{n+1}^{2STA*} \leq \cdot | \mathbf{X}_{1:n} = \mathbf{x}_{1:n}] \\ &= \mathbb{P}[\sigma_{n+1}^2 - \hat{\sigma}_{n+1}^{2STA*} \leq \cdot | \mathbf{X}_{1:n} = \mathbf{x}_{1:n}] * \mathbb{P}[\sigma_{n+1|1:n}^2(\varepsilon_{n+1}^2 - 1) \leq \cdot], \end{aligned} \quad (2.36)$$

where $\hat{\sigma}_{n+1}^{2STA*}$ is defined in (2.18), the desired prediction interval, say J_γ^{STA} , leads to a sensible probabilistic statement due to variability in ε_{n+1}^2 :

$$\mathbb{P}\left[X_{n+1}^2 \in J_\gamma^{STA}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) \middle| \mathbf{X}_{1:n} = \mathbf{x}_{1:n}\right] \underset{(\approx)}{=} 1 - \gamma. \quad (2.37)$$

⁴For independent variables X and Y with $X \sim F_X$, $Y \sim F_Y$ and $Z = X + Y \sim F_Z$, we write $F_Z = F_X * F_Y$ to denote $F_Z(z) = \int_{-\infty}^z F_X(z-y)dF_Y(y)$.

However, it cannot incorporate parameter uncertainty either, since the conditional distribution $\mathbb{P}[\sigma_{n+1}^2 - \hat{\sigma}_{n+1}^{2 STA*} \leq \cdot | \mathbf{X}_{1:n} = \mathbf{x}_{1:n}] = \mathbb{P}[\sigma_{n+1|1:n}^2 - \hat{\sigma}_{n+1|1:n}^{2 STA} \leq \cdot]$ is degenerate.

In his textbook Pesaran (2015) resorts to a Bayesian-akin approach to avoid the fundamental issue in forecasting. He argues that θ , although “fixed at the estimation stage ... is viewed best as a random variable at the forecasting stage” (p. 389). Consequently, he assigns some posterior distribution to θ motivated by an uninformed prior. Treating θ not fixed but random⁵, the fundamental issue does not arise.

Barndorff-Nielsen and Cox (1996) require the existence of a transitive statistic $U = U(\mathbf{X}_{1:n})$ of fixed low dimension to establish conditional independence between the sample $\mathbf{X}_{1:n}$ and their considered future random variable given $U = u$. Vidoni (2004, 2009a, 2009b, 2017), Kabaila (1999), Kabaila and He (2004), and Kabaila and Syuhada (2008, 2010) extend their approach and derive improved prediction intervals. Although these methods absorb an additional $O(n^{-1})$ term in the associated conditional coverage probability, there are several drawbacks associated with them: the innovation distribution needs typically be specified (e.g. Gaussian), the results apply only to a limited set of estimators (e.g. maximum likelihood) and their framework can only incorporate finite autoregressive components (e.g. $\text{AR}(p)$).

Assuming two independent processes with the same stochastic structure, using one for prediction and one for the estimation of the parameters, alleviates the fundamental issue faced in the continued Examples 2.1 and 2.2. As the conditional distributions of the 2IP and SPL estimators merge in probability by Theorem 2.1, the 2IP assumption can be avoided by following a sample-split approach as described in Section 2.2.3.

2.5 Conclusion

In this chapter we study the construction of confidence intervals for conditional objects such as conditional means or conditional variances, focusing on the conceptual issue that arises in the process of taking parameter uncertainty into account.

⁵We abstract here from random coefficient models, in which the parameters are treated as random variables potentially depending on hyperparameters.

It stems from the fact that on one hand one must condition on the sample as the past informs about the present and future, yet on the other hand one must allow the data up to now to be treated as random to account for estimation uncertainty. Assuming two independent processes with the same stochastic structure, where one is used for conditioning and one for the estimation of the parameters, bypasses this issue, but the assumption itself can generally not be justified in applications. To avoid this assumption, we propose a solution based on a simple sample-split approach, that requires a much more realistic weak dependence condition instead. To acknowledge that the conditional quantities vary over time, we employ a merging concept generalizing the notion of weak convergence. The conditional distributions of the sample-split estimator and the estimator based on the two-independent-processes assumption are shown to merge in probability under mild conditions. The corresponding sample-split intervals are shown to coincide asymptotically with the intervals commonly constructed by practitioners, which provides a novel and theoretically satisfactory justification for commonly constructed confidence intervals for conditional objects, applicable to a wide class of time series models, including ARMA and GARCH-type models.

One limitation to our approach is that we restrict ourselves to univariate time series and objects of interests. At the expense of more involved notation this could be extended to multivariate time series and objects of interests. A second, and more restrictive, limitation is our weak dependence assumption needed to achieve asymptotic independence between the two subsamples, which for instance rules out application to integrated processes. Given the fundamental role of this assumption in our setup, it appears difficult to generalize this. However, this also casts further doubt on the two-independent-processes assumption as validation for confidence intervals constructed for such persistent processes. A case-by-case treatment, as for instance done by Gospodinov (2002) for near unit root processes and Samarayake and Hasza (1988) for explosive processes, appears to be necessary in such cases, and standard confidence intervals should be treated with caution.

2.A Lemmas and Proofs of Theorems

We only present the proofs of the leading results here. The proofs of the lemmas and corollaries can be found in Appendix 2.B. Before going to the proofs, we first introduce the following auxiliary metrics that will be encountered in the proofs. For arbitrary cdfs F and G on \mathbb{R} , the Kolmogorov and Lévy metric are

$$d_K(F, G) = \sup_{\tau \in \mathbb{R}} |F(\tau) - G(\tau)|$$

$$d_L(F, G) = \inf \left\{ \zeta > 0 : G(\tau - \zeta) - \zeta \leq F(\tau) \leq G(\tau + \zeta) + \zeta \quad \forall \tau \in \mathbb{R} \right\}.$$

Moreover, let $Z_\infty \sim G_\infty$ (with G_∞ given in Assumption 2.1.(i)) be defined on some probability space $(\check{\Omega}, \check{\mathcal{F}}, \check{\mathbb{P}})$ and define the product measure $\check{\mathbb{P}} = \mathbb{P} \times \check{\mathbb{P}}$ on the space $\Omega \times \check{\Omega}$ with σ -field, generated by the *measurable rectangles* $A \times \check{A}$ with $A \in \mathcal{F}$ and $\check{A} \in \check{\mathcal{F}}$ (cf. Billingsley, 1986, Theorem 18.2). Notice that Z_∞ is independent of $\{X_t\}$ and $\{Y_t\}$ by construction.

In the proofs we often consider the “doubly random” versions of estimators and intervals, where the first function argument, the sample to condition on, is considered random. Instead, we account for the conditioning in the probability statements. This notation is more convenient for proving the results, as we need them to hold for “all sequences $\mathbf{x}_{1:n}$ occurring with high probability”, which is much easier to quantify by treating these sequences as random. Therefore we use notations such as the unconditional estimators $\hat{\psi}_{n+1}^{2IP} = \hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{Y}_{1:n})$ and $\hat{\psi}_{n+1}^{SPL} = \hat{\psi}_{n+1}(\mathbf{X}_{nP:n}^c, \mathbf{X}_{1:nE})$, as well as their corresponding intervals $I^{2IP}(\mathbf{X}_{1:n}, \mathbf{Y}_{1:n})$ and $I^{SPL}(\mathbf{X}_{nP:n}^c, \mathbf{X}_{1:nE})$; also see (2.24),(2.25) and the remarks above Theorem 2.3.

2.A.1 Lemmas

Lemma 2.1. *Let*

$$R_n^{2IP} := m_n \left(\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{Y}_{1:n}) - \psi_{n+1} \right) - \frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}, \theta_0)}{\partial \theta'} m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0) \quad (2.38)$$

$$R_n^{SPL} := m_n \left(\hat{\psi}_{n+1}(\mathbf{X}_{nP:n}^c, \mathbf{X}_{1:nE}) - \psi_{n+1} \right) - \frac{\partial \psi_{n+1}^s(\mathbf{X}_{nP:n}^c, \theta_0)}{\partial \theta'} m_n(\hat{\theta}(\mathbf{X}_{1:nE}) - \theta_0). \quad (2.39)$$

(i) *If Assumptions 2.1.(i), 2.1.(ii), 2.1.(iv), 2.1.(v) and 2.2.(i) hold, then R_n^{2IP}*

is $o_p(1)$;

(ii) if Assumptions 2.1.(i), 2.1.(ii), 2.1.(iv), 2.1.(v) and 2.3.(i) hold, then R_n^{SPL} is $o_p(1)$.

Lemma 2.2. Let $G_{n_E}^{SPL}(\cdot|\mathcal{I}_{n_P:n})$ be as given in Assumption 2.3.(iii) and denote the conditional cdf of $Z_n^{2IP} := m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)$ given $\mathcal{I}_{1:n}$ by $G_n^{2IP}(\cdot|\mathcal{I}_{1:n})$.

(i) Under Assumptions 2.1.(i) and 2.2, $\int h dG_n^{2IP}(\cdot|\mathcal{I}_{1:n}) \xrightarrow{P} \int h dG_\infty \forall h \in BL$;

(ii) Under Assumptions 2.1.(i) and 2.3, $\int h dG_{n_E}^{SPL}(\cdot|\mathcal{I}_{n_P:n}) \xrightarrow{P} \int h dG_\infty \forall h \in BL$.

Lemma 2.3. Let $\{G_n\}$ be a sequence of cdfs and G be a (non-random) cdf on $(\mathbb{R}^r, \|\cdot\|)$. If $\int h dG_n \xrightarrow{P} \int h dG$ for all $h \in BL$, then $\sup_{h \in \mathcal{H}} \left| \int h d(G_n - G) \right| \xrightarrow{P} 0$, where $\mathcal{H} = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} \leq 1\}$.

Lemma 2.4. Assume that the \mathbb{R}^r -valued random variable w_n (row vector) is $O_p(1)$ and \mathcal{I}_n -measurable. Further, suppose the real-valued random variable R_n is $o_p(1)$ and the \mathbb{R}^r -valued random variable Z_n satisfies $\mathbb{P}[Z_n \leq \cdot|\mathcal{I}_n] \xrightarrow{P} G_\infty$. Then, the two sequences of conditional cdfs $\mathbb{P}[w_n Z_n + R_n \leq \cdot|\mathcal{I}_n]$ and $\mathbb{P}[w_n Z_\infty \leq \cdot|\mathcal{I}_n]$ merge in probability.

Lemma 2.5. Let $\epsilon > 0$ and F and G be cdfs on \mathbb{R} with $G(\tau - \epsilon) - \epsilon \leq F(\tau) \leq G(\tau + \epsilon) + \epsilon$ for all $\tau \in \mathbb{R}$. Then $F^{-1}(u - \epsilon) - \epsilon \leq G^{-1}(u) \leq F^{-1}(u + \epsilon) + \epsilon$ for all $u \in (\epsilon, 1 - \epsilon)$.

Lemma 2.6. Suppose $\{F_n\}$ and $\{G_n\}$ are sequences of conditional cdfs with $d_L(F_n, G_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Further, assume that G_n is stochastically uniformly equicontinuous: for every $\epsilon, \eta > 0$, there exist $\delta = \delta(\epsilon, \eta) > 0$ and $\bar{n} = \bar{n}(\epsilon, \eta)$ such that $\mathbb{P}[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau - \tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon] < \eta$ for all $n \geq \bar{n}$. Then, $d_K(F_n, G_n) \xrightarrow{P} 0$.

Lemma 2.7. If the sequences of conditional cdfs $\{F_n\}$ and $\{G_n\}$ merge in probability and G_n is stochastically uniformly equicontinuous, then $\mathbb{P}[G_n^{-1}(\gamma_1) \leq M_n \leq G_n^{-1}(1 - \gamma_2)|\mathcal{I}_n] \xrightarrow{P} 1 - \gamma_1 - \gamma_2$ whenever $0 \leq \gamma_1 \leq 1 - \gamma_2 \leq 1$, where the random variable M_n given the σ -algebra \mathcal{I}_n has the cdf F_n .

Lemma 2.8. Suppose $\{F_n\}$ and $\{G_n\}$ are sequences of conditional cdfs with $d_L(F_n, G_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Further, assume that G_n^{-1} is stochastically pointwise equicontinuous at $u \in (0, 1)$: for all $\epsilon, \eta > 0$, there exist $\delta = \delta(\epsilon, \eta, u) > 0$ and $\bar{n} = \bar{n}(\epsilon, \eta, u)$ such that $\mathbb{P}[\sup_{|u-v| < \delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \epsilon] < \eta$ for all $n \geq \bar{n}$. Then $|F_n^{-1}(u) - G_n^{-1}(u)| \xrightarrow{P} 0$.

2.A.2 Proofs of Theorems.

Proof of Theorem 2.1. Let w_n^{2IP} equal the transpose of $\frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta_0)}{\partial \theta}$ and w_n^{SPL} the transpose of $\frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta}$ and set

$$F_{\infty,n}^{2IP}(\tau|\mathcal{I}_{1:n}) := \bar{\mathbb{P}}[w_n^{2IP} Z_\infty \leq \tau|\mathcal{I}_{1:n}], \quad (2.40)$$

$$F_{\infty,n}^{SPL}(\tau|\mathcal{I}_{n_P:n}) := \bar{\mathbb{P}}[w_n^{SPL} Z_\infty \leq \tau|\mathcal{I}_{n_P:n}], \quad (2.41)$$

the conditional cdfs of $w_n^{2IP} Z_\infty$ given $\mathcal{I}_{1:n}$ and $w_n^{SPL} Z_\infty$ given $\mathcal{I}_{n_P:n}$, respectively. Note that (2.40) and (2.41) can be considered to be the “merging limits” of $w_n^{2IP} m_n(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0)$ and $w_n^{SPL} m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$. Since $m_n(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0)$ and $m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$ converge in distribution to G_∞ but w_n^{2IP} and w_n^{SPL} do not converge, we indexed the “merging limits” in (2.40) and (2.41) by ∞ and n . The triangle inequality implies

$$\begin{aligned} & d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), F_n^{SPL}(\cdot|\mathcal{I}_{n_P:n})\right) \\ & \leq \underbrace{d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})\right)}_I + \underbrace{d_{BL}\left(F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n}), F_{\infty,n}^{SPL}(\cdot|\mathcal{I}_{n_P:n})\right)}_{II} \\ & \quad + \underbrace{d_{BL}\left(F_{\infty,n}^{SPL}(\cdot|\mathcal{I}_{n_P:n}), F_n^{SPL}(\cdot|\mathcal{I}_{n_P:n})\right)}_{III}. \end{aligned}$$

We prove that I , II and III are $o_p(1)$. Consider I and note that $m_n(\hat{\psi}_{n+1}^{2IP} - \psi_{n+1}) = w_n^{2IP} Z_n^{2IP} + R_n^{2IP}$, where $Z_n^{2IP} = m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)$. The weight w_n^{2IP} is $\mathcal{I}_{1:n}$ measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v), R_n^{2IP} is $o_p(1)$ by Lemma 2.1 and $\int h dG_n^{2IP}(\cdot|\mathcal{I}_{1:n}) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$ by Lemma 2.2. Replacing Z_n , R_n , w_n and \mathcal{I}_n in Lemma 2.4 by Z_n^{2IP} , R_n^{2IP} , w_n^{2IP} and $\mathcal{I}_{1:n}$ implies that $F_n^{2IP}(\tau|\mathcal{I}_{1:n}) = \bar{\mathbb{P}}[w_n^{2IP} Z_n^{2IP} + R_n^{2IP} \leq \tau|\mathcal{I}_{1:n}]$ and $F_{\infty,n}^{2IP}(\tau|\mathcal{I}_{1:n}) = \bar{\mathbb{P}}[w_n^{2IP} Z_\infty \leq \tau|\mathcal{I}_{1:n}]$ merge in probability, i.e. $I \xrightarrow{P} 0$.

Consider II and rewrite $w_n^{SPL} Z_\infty = w_n^{2IP} Z_\infty + (w_n^{SPL} - w_n^{2IP}) Z_\infty$. The weight w_n^{2IP} is $\mathcal{I}_{1:n}$ measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v), $(w_n^{SPL} - w_n^{2IP}) Z_\infty = o_p(1) O_p(1) = o_p(1)$ by Assumption 2.1.(v) and $\int h d\bar{\mathbb{P}}[Z_\infty \leq \cdot|\mathcal{I}_{1:n}] = \int h dG_\infty$ for each $h \in BL$. Replacing Z_n , R_n , w_n and \mathcal{I}_n in Lemma 2.4 by Z_∞ , $(w_n^{SPL} - w_n^{2IP}) Z_\infty$, w_n^{2IP} and $\mathcal{I}_{1:n}$ implies that $F_{\infty,n}^{2IP}(\tau|\mathcal{I}_{1:n}) = \bar{\mathbb{P}}[w_n^{2IP} Z_\infty \leq \tau|\mathcal{I}_{1:n}]$ and $F_{\infty,n}^{SPL}(\tau|\mathcal{I}_{n_P:n}) = \bar{\mathbb{P}}[w_n^{SPL} Z_\infty \leq \tau|\mathcal{I}_{1:n}]$ merge in probability, i.e. $II \xrightarrow{P} 0$.

Consider III and note that $m_n(\hat{\psi}_{n+1}^{SPL} - \psi_{n+1}) = w_n^{SPL} Z_n^{SPL} + R_n^{SPL} +$

S_n^{SPL} , where $Z_n^{SPL} = m_{n_E}(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)$ and $S_n^{SPL} = (\frac{m_n}{m_{n_E}} - 1)w_n^{SPL}Z_n^{SPL}$. The weight w_n^{SPL} is $\mathcal{I}_{n_P:n}$ measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v), R_n^{SPL} is $o_p(1)$ by Lemma 2.1 and $\int h dG_{n_E}^{SPL}(\cdot|\mathcal{I}_{n_P:n}) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$ by Lemma 2.2. Further, S_n^{SPL} is $o_p(1)$ since $w_n^{SPL} = O_p(1)$ and $(\frac{m_n}{m_{n_E}} - 1)Z_n^{SPL} = o_p(1)$ by Assumptions 2.1.(i) and 2.3.(i). Replacing Z_n , R_n , w_n and \mathcal{I}_n in Lemma 2.4 by Z_n^{SPL} , $R_n^{SPL} + S_n^{SPL}$, w_n^{SPL} and $\mathcal{I}_{n_P:n}$ implies that $F_n^{SPL}(\tau|\mathcal{I}_{n_P:n}) = \bar{\mathbb{P}}[w_n^{SPL}Z_n^{SPL} + R_n^{SPL} + S_n^{SPL} \leq \tau|\mathcal{I}_{n_P:n}]$ and $F_{\infty,n}^{SPL}(\tau|\mathcal{I}_{n_P:n}) = \bar{\mathbb{P}}[w_n^{SPL}Z_\infty \leq \tau|\mathcal{I}_{n_P:n}]$ merge in probability, i.e. $III \xrightarrow{P} 0$. \square

Proof of Theorem 2.2. Consider statement 1.(i). Let \hat{Z}_n^{2IP} follow the mixture distribution $\widehat{G}_n(\cdot)$ as a function of $\mathbf{Y}_{1:n}$ such that given $\mathbf{Y}_{1:n}$ the conditional distribution of the random variable \hat{Z}_n^{2IP} is $\widehat{G}_n(\cdot)$. Further, let

$$\widehat{F}_n^{2IP}(\cdot) \text{ be the conditional cdf of } \hat{w}_n^{2IP}\hat{Z}_n^{2IP} \text{ given } \mathcal{H}_{1:n} \quad (2.42)$$

where \hat{w}_n^{2IP} equals $\frac{\partial \psi_n^s(\mathbf{X}_{1:n}; \hat{\theta}(\mathbf{Y}_{1:n}))}{\partial \theta'}$ and $\mathcal{H}_{1:n} = \sigma(X_1, \dots, X_n, Y_1, \dots, Y_n)$. Then

$$\begin{aligned} & d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), \widehat{F}_n^{2IP}(\cdot)\right) \\ & \leq d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})\right) + d_{BL}\left(F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n}), \widehat{F}_n^{2IP}(\cdot)\right) \end{aligned}$$

by the triangle inequality, where $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ is defined in equation (2.40). In the proof of Theorem 2.1, we have shown that $F_n^{2IP}(\cdot|\mathcal{I}_{1:n})$ and $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ merge in probability under Assumptions 2.1 and 2.2. It suffices to show that $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ and $\widehat{F}_n^{2IP}(\cdot)$ merge in probability. Write $\hat{w}_n^{2IP}\hat{Z}_n^{2IP} = w_n^{2IP}\hat{Z}_n^{2IP} + \hat{R}_n^{2IP}$ with $\hat{R}_n^{2IP} = (\hat{w}_n^{2IP} - w_n^{2IP})\hat{Z}_n^{2IP}$. First, we show $\hat{R}_n^{2IP} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} & \mathbb{P}\left[|\hat{R}_n^{2IP}| \geq \epsilon\right] \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{2IP}\| \geq \epsilon\right] \\ & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{2IP}\| \geq \epsilon \cap \dot{\theta}_n \in \mathcal{V}(\theta_0)\right] \\ & \quad + \mathbb{P}\left[\dot{\theta}_n \notin \mathcal{V}(\theta_0)\right] \\ & \leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{2IP}\| \geq \epsilon\right] \\ & \quad + \mathbb{P}\left[\dot{\theta}_n \notin \mathcal{V}(\theta_0)\right], \end{aligned}$$

where $\hat{\theta}_n$ lies between $\hat{\theta}(\mathbf{Y}_{1:n})$ and θ_0 . The first term of the sum vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v), $\|\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0\| = O_p(m_n^{-1})$ by Assumptions 2.1.(i) and 2.2.(i) and $\|\hat{Z}_n^{2IP}\| = O_p(1)$ as $\hat{Z}_n^{2IP} \sim \widehat{G}_n(\cdot)$ and $\int h d\widehat{G}_n \xrightarrow{P} \int h dG_\infty$ for all $h \in BL$ by Assumptions 2.2.(i) and 2.4. Further, as $\hat{\theta}(\mathbf{Y}_{1:n}) \xrightarrow{P} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\hat{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_n^{2IP} = o_p(1)$ follows. Moreover, w_n^{2IP} is \mathcal{H}_n -measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v) and $\int h d\mathbb{P}[\hat{Z}_n^{2IP} \leq \cdot | \mathcal{H}_{1:n}] = \int h d\widehat{G}_n \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$. Replacing Z_n , R_n , w_n and \mathcal{I}_n in Lemma 2.4 by \hat{Z}_n^{2IP} , \hat{R}_n^{2IP} , w_n^{2IP} and \mathcal{H}_n implies that $\widehat{F}_n^{2IP}(\cdot) = \mathbb{P}[w_n^{2IP} \hat{Z}_n^{2IP} + \hat{R}_n^{2IP} \leq \cdot | \mathcal{H}_{1:n}]$ and $\mathbb{P}[w_n^{2IP} Z_\infty \leq \cdot | \mathcal{H}_{1:n}] = F_{\infty,n}^{2IP}(\cdot | \mathcal{I}_{1:n})$ merge in probability.

Consider statement 1.(ii). As $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$ and $\widehat{F}_n^{2IP}(\cdot)$ merge in probability and $\widehat{F}_n^{2IP}(\cdot)$ is assumed to be stochastically uniformly continuous, Lemma 2.7 applies. Replacing F_n , G_n , M_n and \mathcal{I}_n by $F_n^{2IP}(\cdot | \mathcal{I}_{1:n})$, $\widehat{F}_n^{2IP}(\cdot)$, $m_n(\hat{\psi}_{n+1}^{2IP} - \psi_{n+1})$ and $\mathcal{I}_{1:n}$, respectively, it follows that

$$\begin{aligned} & \mathbb{P}\left[I_\gamma^{2IP}(\mathbf{X}_{1:n}, \mathbf{Y}_{1:n}) \ni \psi_{n+1} \middle| \mathcal{I}_{1:n}\right] \\ &= \mathbb{P}\left[\widehat{F}_n^{2IP}{}^{-1}(\gamma_1) \leq m_n(\hat{\psi}_{n+1}^{2IP} - \psi_{n+1}) \leq \widehat{F}_n^{2IP}{}^{-1}(1 - \gamma_2) \middle| \mathcal{I}_{1:n}\right] \xrightarrow{P} 1 - \gamma. \end{aligned}$$

Claim 2.(i) is similarly proven as 1.(i). Let \hat{Z}_n^{SPL} follow the mixture distribution $\widehat{G}_n(\cdot)$ as a function of $\mathbf{X}_{1:n_E}$ such that given $\mathbf{X}_{1:n_E}$ the conditional distribution of the random variable \hat{Z}_n^{SPL} is $\widehat{G}_n(\cdot)$. Further, let

$$\widehat{F}_n^{SPL}(\cdot) \text{ be the conditional cdf of } \hat{w}_n^{SPL} \hat{Z}_n^{SPL} \text{ given } \mathcal{I}_{1:n} \quad (2.43)$$

where \hat{w}_n^{SPL} equals the transpose of $\frac{\partial \psi_n^s(\mathbf{X}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E}))}{\partial \theta}$. Then

$$\begin{aligned} & d_{BL}\left(F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n}), \widehat{F}_n^{SPL}(\cdot)\right) \\ & \leq d_{BL}\left(F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n}), F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n})\right) + d_{BL}\left(F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n}), \widehat{F}_n^{SPL}(\cdot)\right), \end{aligned}$$

where $F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ is defined in equation (2.41). In the proof of Theorem 2.1, we have shown that $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ and $F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ merge in probability under Assumptions 2.1 and 2.3. It suffices to show that $F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ and $\widehat{F}_n^{SPL}(\cdot)$ merge in probability. Write $\hat{w}_n^{SPL} \hat{Z}_n^{SPL} = w_n^{SPL} \hat{Z}_n^{SPL} + \hat{R}_n^{SPL}$ with $\hat{R}_n^{SPL} = (\hat{w}_n^{SPL} - w_n^{SPL}) \hat{Z}_n^{SPL}$. First, we show $\hat{R}_n^{SPL} = o_p(1)$. Take an arbitrary $\epsilon > 0$.

We obtain

$$\begin{aligned}
 \mathbb{P}\left[|\hat{R}_n^{SPL}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\right\| \|\hat{Z}_n^{SPL}\| \geq \epsilon\right] \\
 &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\right\| \|\hat{Z}_n^{SPL}\| \geq \epsilon \cap \ddot{\theta}_n \in \mathcal{V}(\theta_0)\right] \\
 &\quad + \mathbb{P}\left[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)\right] \\
 &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\right\| \|\hat{Z}_n^{SPL}\| \geq \epsilon\right] \\
 &\quad + \mathbb{P}\left[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)\right],
 \end{aligned}$$

where $\ddot{\theta}_n$ lies between $\hat{\theta}(\mathbf{X}_{1:n_E})$ and θ_0 . The first term of the sum vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'}\right\|$ is $O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v) and $\left\|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\right\| = O_p(m_{n_E}^{-1})$ by Assumptions 2.1.(i) and 2.3.(i), $\|\hat{Z}_n^{SPL}\| = O_p(1)$ as $\hat{Z}_n^{SPL} \sim \widehat{G}_n(\cdot)$ and $\int h d\widehat{G}_n \xrightarrow{P} \int h dG_\infty$ for all $h \in BL$ by Assumptions 2.3.(i), 2.3.(ii) and 2.4. Further, as $\hat{\theta}(\mathbf{X}_{1:n_E}) \xrightarrow{P} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_n^{SPL} = o_p(1)$ follows. Moreover, w_n^{SPL} is $\mathcal{I}_{1:n}$ -measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v) and $\int h d\mathbb{P}[\hat{Z}_n^{SPL} \leq \cdot | \mathcal{I}_{1:n}] = \int h d\widehat{G}_n(\cdot) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$. Replacing Z_n, R_n, w_n and \mathcal{I}_n in Lemma 2.4 by $\hat{Z}_n^{SPL}, \hat{R}_n^{SPL}, w_n^{SPL}$ and $\mathcal{I}_{1:n}$ implies that $\widehat{F}_n^{SPL}(\cdot) = \mathbb{P}[w_n^{SPL} \hat{Z}_n^{SPL} + \hat{R}_n^{SPL} \leq \cdot | \mathcal{I}_{1:n}]$ and $\mathbb{P}[w_n^{SPL} Z_\infty \leq \cdot | \mathcal{I}_{1:n}] = F_{\infty,n}^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ merge in probability.

The proof of statement 2.(ii) is similar to the proof of claim 1.(ii). As $\widehat{F}_n^{SPL}(\cdot)$ and $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})$ merge in probability and $\widehat{F}_n^{SPL}(\cdot)$ is assumed to be stochastically uniformly continuous, Lemma 2.7 applies. Replacing F_n, G_n, M_n and \mathcal{I}_n by $F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n}), \widehat{F}_n^{SPL}(\cdot), m_n(\hat{\psi}_{n+1}^{SPL} - \psi_{n+1})$ and $\mathcal{I}_{n_P:n}$, respectively, yields

$$\begin{aligned}
 &\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{X}_{n_P:n}, \mathbf{X}_{1:n_E}) \ni \psi_{n+1} | \mathcal{I}_{n_P:n}\right] \\
 &= \mathbb{P}\left[\widehat{F}_n^{SPL}{}^{-1}(\gamma_1) \leq m_n(\hat{\psi}_{n+1}^{SPL} - \psi_{n+1}) \leq \widehat{F}_n^{SPL}{}^{-1}(1 - \gamma_2) | \mathcal{I}_{n_P:n}\right] \xrightarrow{P} 1 - \gamma,
 \end{aligned}$$

which completes the proof. \square

Proof of Theorem 2.3. Consider the first statement and expand

$$\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) - \hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E})$$

$$= (\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) - \psi_{n+1}) - (\hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E}) - \psi_{n+1}).$$

We show that both terms are $o_p(1)$. Using (2.39), we have

$$\begin{aligned} & \hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E}) - \psi_{n+1} \\ &= \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0) + m_n^{-1} R_n^{SPL}, \end{aligned}$$

where $\frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta'} = O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v) and $\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0 = o_p(1)$ by Assumptions 2.1.(i), 2.3.(i) and 2.3.(ii). Together with $R_n^{SPL} = o_p(1)$ by Lemma 2.1 and $m_n^{-1} = o(1)$, it implies that $\hat{\psi}_{n+1}(\mathbf{X}_{n_P:n}^c, \mathbf{X}_{1:n_E}) - \psi_{n+1} = o_p(1)$. In addition, replacing $\mathbf{Y}_{1:n}$ by $\mathbf{X}_{1:n}$ in equation (2.38), we get

$$\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) - \psi_{n+1} = \frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}, \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0) + m_n^{-1} R_n^{STA},$$

where R_n^{STA} is obtained by replacing $\mathbf{Y}_{1:n}$ by $\mathbf{X}_{1:n}$ in Equation (2.38). We have $\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}, \theta_0)}{\partial \theta} \right\| = O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v) and $\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0 = o_p(1)$ by Assumption 2.1.(i). Since $R_n^{2IP} = o_p(1)$ has been shown in Lemma 2.1 without using Assumption 2.2.(ii), we have $R_n^{STA} = o_p(1)$. Together with $m_n^{-1} = o(1)$, it follows that $\hat{\psi}_{n+1}(\mathbf{X}_{1:n}, \mathbf{X}_{1:n}) - \psi_{n+1} = o_p(1)$ completing the claim.

Consider the second statement and let \hat{Z}_n^{STA} follow the mixture distribution $\hat{G}_n(\cdot)$ as a function of $\mathbf{X}_{1:n}$ such that given $\mathbf{X}_{1:n}$ the conditional cdf of the random variable \hat{Z}_n^{STA} is $\hat{G}_n(\cdot)$. Further, let

$$\widehat{F}_n^{STA}(\cdot) \text{ be the conditional cdf of } \hat{w}_n^{STA} \hat{Z}_n^{STA} \text{ given } \mathcal{I}_{1:n} \quad (2.44)$$

where \hat{w}_n^{STA} equals the transpose of $\frac{\partial \psi_n^s(\mathbf{X}_{1:n}; \hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta}$. First, we show that $\widehat{F}_n^{STA}(\cdot)$ and $\widehat{F}_n^{SPL}(\cdot)$, defined in (2.43), merge in probability. The triangle inequality implies

$$\begin{aligned} & d_{BL}(\widehat{F}_n^{SPL}(\cdot), \widehat{F}_n^{STA}(\cdot)) \\ & \leq d_{BL}(\widehat{F}_n^{SPL}(\cdot), F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n})) + d_{BL}(F_n^{SPL}(\cdot | \mathcal{I}_{n_P:n}), F_n^{2IP}(\cdot | \mathcal{I}_{1:n})) \\ & \quad + d_{BL}(F_n^{2IP}(\cdot | \mathcal{I}_{1:n}), \widehat{F}_n^{STA}(\cdot)), \end{aligned}$$

where the first two terms on the right hand side converge in probability to zero

by Theorem 2.2 (statement 2.(i)) and Theorem 2.1, respectively. We are left to show that $F_n^{2IP}(\cdot|\mathcal{I}_{1:n})$ and $\widehat{F_n^{STA}}(\cdot)$ merge in probability. The triangle inequality implies that

$$\begin{aligned} & d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), \widehat{F_n^{STA}}(\cdot)\right) \\ & \leq d_{BL}\left(F_n^{2IP}(\cdot|\mathcal{I}_{1:n}), F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})\right) + d_{BL}\left(F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n}), \widehat{F_n^{STA}}(\cdot)\right), \end{aligned}$$

where $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ is defined in equation (2.40). In the proof of Theorem 2.1, we have shown that $F_n^{2IP}(\cdot|\mathcal{I}_{1:n})$ and $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ merge in probability under Assumptions 2.1 and 2.2. It suffices to show that $F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ and $\widehat{F_n^{STA}}(\cdot)$ merge in probability. Write $\hat{w}_n^{STA} \hat{Z}_n^{STA} = w_n^{2IP} \hat{Z}_n^{STA} + \hat{R}_n^{STA}$ with $\hat{R}_n^{STA} = (\hat{w}_n^{STA} - w_n^{2IP}) \hat{Z}_n^{STA}$ (note that, in contrast to \hat{w}_n^{STA} , there is no need to introduce w_n^{STA} as it equals w_n^{2IP}). First, we show $\hat{R}_n^{STA} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} & \mathbb{P}\left[|\hat{R}_n^{STA}| \geq \epsilon\right] \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{STA}\| \geq \epsilon\right] \\ & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{STA}\| \geq \epsilon \cap \ddot{\theta}_n \in \mathcal{V}(\theta_0)\right] \\ & \quad + \mathbb{P}\left[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)\right] \\ & \leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0\right\| \|\hat{Z}_n^{STA}\| \geq \epsilon\right] \\ & \quad + \mathbb{P}\left[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)\right]. \end{aligned}$$

where $\ddot{\theta}_n$ lies between $\hat{\theta}(\mathbf{X}_{1:n})$ and θ_0 . The first term of the sum vanishes since $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'}\right\|$ is $O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v), $\left\|\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0\right\| = O_p(m_n^{-1})$ by Assumption 2.1.(i) and $\|\hat{Z}_n^{STA}\| = O_p(1)$ as $\hat{Z}_n^{STA} \sim \widehat{G}_n(\cdot) \xrightarrow{p} G_\infty$ by Assumption 2.4. Further, as $\hat{\theta}(\mathbf{X}_{1:n}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_n^{STA} = o_p(1)$ follows. Moreover, w_n^{2IP} is $\mathcal{I}_{1:n}$ -measurable and $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v) and $\int h d\mathbb{P}[\hat{Z}_n^{STA} \leq \cdot|\mathcal{I}_{1:n}] = \int h d\widehat{G}_n(\cdot) \xrightarrow{p} \int h dG_\infty$ for each $h \in BL$. Replacing Z_n , R_n , w_n and \mathcal{I}_n in Lemma 2.4 by \hat{Z}_n^{STA} , \hat{R}_n^{STA} , w_n^{2IP} and $\mathcal{I}_{1:n}$ implies that $\widehat{F_n^{STA}}(\cdot) = \mathbb{P}[w_n^{2IP} \hat{Z}_n^{STA} + \hat{R}_n^{STA} \leq \cdot|\mathcal{I}_{1:n}]$ and $\mathbb{P}[w_n^{2IP} Z_\infty \leq \cdot|\mathcal{I}_{1:n}] = F_{\infty,n}^{2IP}(\cdot|\mathcal{I}_{1:n})$ merge in probability. Thus, $\widehat{F_n^{SPL}}(\cdot)$ and $\widehat{F_n^{STA}}(\cdot)$ merge in probability. Together with $\widehat{F_n^{SPL}}(\cdot)$ being stochastically pointwise continuous at γ_1 and $1 - \gamma_2$, assertion

(2.31) follows by Lemma 2.8, which completes the proof. \square

2.B Additional Proofs

2.B.1 Proofs of Lemmas

Proof of Lemma 2.1. Consider (i). By Assumption 2.1.(ii) one can write R_n^{2IP} as follows:

$$\begin{aligned} R_n^{2IP} = & \underbrace{m_n(\psi_{n+1}^s(\mathbf{X}_{1:n}; \theta_0) - \psi_{n+1}(X_n, X_{n-1}, \dots; \theta_0))}_{=R_{1,n}^{2IP}} \\ & + \underbrace{(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)' \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \dot{\theta}_n)}{\partial \theta \partial \theta'} m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)}_{=R_{2,n}^{2IP}}, \end{aligned}$$

where $\dot{\theta}_n$ lies between θ_0 and $\hat{\theta}(\mathbf{Y}_{1:n})$. By Assumption 2.1.(v), $R_{1,n}^{2IP}$ is $o_p(1)$; hence we are left to show that $R_{2,n}^{2IP} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} \mathbb{P}[|R_{2,n}^{2IP}| \geq \epsilon] & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)\right\|^2 \geq \epsilon\right] \\ & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)\right\|^2 \geq \epsilon \cap \dot{\theta}_n \in \mathcal{V}(\theta_0)\right] \\ & \quad + \mathbb{P}[\dot{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'}\right\| \left\|m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)\right\|^2 \geq \epsilon\right] \\ & \quad + \mathbb{P}[\dot{\theta}_n \notin \mathcal{V}(\theta_0)]. \end{aligned}$$

The first term vanishes since $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta)}{\partial \theta \partial \theta'}\right\| = O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v) and $m_n \|\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0\|^2 = O_p(m_n^{-1})$ by Assumptions 2.1.(i) and 2.2.(i). Further, as $\hat{\theta}(\mathbf{Y}_{1:n}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\dot{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $R_{2,n}^{2IP} = o_p(1)$ follows.

The proof of (ii) is analogous; by Assumption 2.1.(ii) one can express R_n^{SPL} as

follows:

$$\begin{aligned}
 R_n^{SPL} &= \underbrace{m_n(\psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0) - \psi_{n+1}(X_n, X_{n-1}, \dots; \theta_0))}_{=R_{1,n}^{SPL}} \\
 &\quad + \underbrace{(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)' \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \ddot{\theta}_n)}{\partial \theta \partial \theta'} m_n(\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0)}_{=R_{2,n}^{SPL}}
 \end{aligned}$$

with $R_{1,n}^{SPL} = o_p(1)$ by Assumption 2.1.(v) and $\ddot{\theta}_n$ lying between θ_0 and $\hat{\theta}(\mathbf{X}_{1:n_E})$. For an arbitrary $\epsilon > 0$, we obtain

$$\begin{aligned}
 \mathbb{P}[|R_{2,n}^{SPL}| \geq \epsilon] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\|^2 \geq \epsilon\right] \\
 &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\|^2 \geq \epsilon \cap \ddot{\theta}_n \in \mathcal{V}(\theta_0)\right] \\
 &\quad + \mathbb{P}[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)] \\
 &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\|^2 \geq \epsilon\right] \\
 &\quad + \mathbb{P}[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)].
 \end{aligned}$$

The first term vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'}\right\| = O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v) and $m_n \|\hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0\|^2 = O_p(m_n^{-1})$ by Assumptions 2.1.(i) and 2.3.(i). Further, as $\hat{\theta}(\mathbf{X}_{1:n_E}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\ddot{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $R_{2,n}^{SPL} = o_p(1)$ follows. \square

Proof of Lemma 2.2. Consider (i) and let G_n^{2IP} denote the unconditional distribution of $m_n(\hat{\theta}(\mathbf{Y}_{1:n}) - \theta_0)$. By Assumption 2.2.(ii), we have for each $h \in BL$

$$\int h dG_n^{2IP}(\cdot | \mathcal{I}_{1:n}) = \int h dG_n^{2IP} \rightarrow \int h dG_\infty,$$

where the last assertion comes from Assumptions 2.1.(i) and 2.2.(i) and Portman-teau's Lemma (cf. van der Vaart, 2000, Lemma 2.2). Consider (ii); for each $h \in BL$ we obtain

$$\int h d(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_\infty)$$

$$= \underbrace{\int h d(G_{n_E}^{SPL} - G_\infty)}_I + \underbrace{\int h d(G_{n_E}^{SPL}(\cdot | \mathcal{I}_{n_P:n}) - G_{n_E}^{SPL})}_{II},$$

where $I \rightarrow 0$ by Assumptions 2.1.(i), 2.3.(i) and 2.3.(ii) and Portmanteau's Lemma and $II \xrightarrow{P} 0$ by Assumption 2.3.(iii). \square

Proof of Lemma 2.3. For $r = 1$ Lemma 2.3 appears as Lemma 2 of the supplemental material to Castillo and Rousseau (2015). Extending their result to $r > 1$ we closely follow the proof of Dudley (2002, Theorem 11.3.3) and write Q_n and Q to denote the probability measures corresponding to G_n and G , respectively. Let $\epsilon > 0$ and take a compact set $K \subset \mathbb{R}^r$ such that $Q(K) > 1 - \epsilon$. The set of functions $h \in \mathcal{H}$, restricted to K , form a compact set of functions for the supremum norm by the Arzela-Ascoli theorem (cf. Dudley, 2002, Theorem 2.4.7). Thus for some finite $J = J(\epsilon)$ there are $h_1, \dots, h_J \in \mathcal{H}$ such that for any $h \in \mathcal{H}$, there is a $j \leq J$ with $\sup_{y \in K} |h(y) - h_j(y)| < \epsilon$. Let $K^\epsilon = \{y \in \mathbb{R}^r : \|x - y\| < \epsilon \text{ for some } x \in K\}$. One has $\sup_{x \in K^\epsilon} |h(x) - h_j(x)| < 3\epsilon$, since if $y \in K$ and $\|x - y\| < \epsilon$, then

$$\begin{aligned} |h(x) - h_j(x)| &\leq |h(x) - h(y)| + |h(y) - h_j(y)| + |h_j(y) - h_j(x)| \\ &\leq \|h\|_{BL} \|x - y\| + \epsilon + \|h_j\|_{BL} \|x - y\| < 3\epsilon. \end{aligned}$$

Let $g(x) = \max\{0, 1 - \|x - K\|/\epsilon\}$, where $\|x - K\| = \inf\{\|x - y\| : y \in K\}$ for all $x \in \mathbb{R}^r$. Then $g \in BL$ and $I\{x \in K\} \leq g \leq I\{x \in K^\epsilon\}$, where $I\{\cdot\}$ denotes the indicator function. It follows that

$$Q_n(\mathbb{R}^r \setminus K^\epsilon) = 1 - Q_n(K^\epsilon) \leq 1 - \int g dQ_n \xrightarrow{P} 1 - \int g dQ \leq 1 - Q(K) < \epsilon$$

or equivalently $\mathbb{P}[Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon] \rightarrow 0$. Thus, for each $h \in \mathcal{H}$ and h_j as above

$$\begin{aligned} &\sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| \\ &\leq \sup_{h \in \mathcal{H}} \int |h - h_j| d(Q_n + Q) + \left| \int h_j d(Q_n - Q) \right| \\ &\leq 2(Q_n + Q)(\mathbb{R}^r \setminus K^\epsilon) + 6\epsilon + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \\ &\leq 8\epsilon + 2Q_n(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right|. \end{aligned}$$

Hence,

$$\begin{aligned}
 & \mathbb{P} \left[\sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| \geq 11\epsilon \right] \\
 & \leq \mathbb{P} \left[2Q_n(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \geq 3\epsilon \right] \\
 & \leq \mathbb{P} \left[Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \mathbb{P} \left[\max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \geq \epsilon \right] \\
 & \leq \mathbb{P} \left[Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \sum_{j=1}^J \mathbb{P} \left[\left| \int h_j d(Q_n - Q) \right| \geq \epsilon \right],
 \end{aligned}$$

where the last two terms are converging to 0 for finite J noting that $\int h_j d(Q_n - Q) = \int h_j d(G_n - G) \xrightarrow{P} 0$. Observing that $\sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| = \sup_{h \in \mathcal{H}} \left| \int h d(G_n - G) \right|$ completes the proof. \square

Proof of Lemma 2.4. This lemma is related to Lemma 8 in Belyaev and Sjöstedt-De Luna (2000) where the quantity corresponding to $\bar{\mathbb{P}}[w_n Z_\infty \leq \cdot | \mathcal{I}_n]$ is non-random. Set $\mathcal{F} = \{f : \mathbb{R} \rightarrow \mathbb{R} : \|f\|_{BL} \leq 1\}$. The triangle inequality implies

$$\begin{aligned}
 & \sup_{f \in \mathcal{F}} \left| \int [f(w_n Z_n + R_n) - f(w_n Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right| \\
 & \leq \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w_n Z_n + R_n) - f(w_n Z_n)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right|}_{=I} \\
 & \quad + \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w_n Z_n) - f(w_n Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right|}_{=II}.
 \end{aligned}$$

We show that $I \xrightarrow{P} 0$ and $II \xrightarrow{P} 0$. Let $\epsilon > 0$; as $\|f\|_{BL} \leq 1$ for all $f \in \mathcal{F}$ we obtain

$$\begin{aligned}
 I & \leq \sup_{f \in \mathcal{F}} \int \left| f(w_n Z_n + R_n) - f(w_n Z_n) \right| d\mathbb{P}[\cdot | \mathcal{I}_n] \\
 & = \sup_{f \in \mathcal{F}} \int_{|R_n| \leq \epsilon} \left| f(w_n Z_n + R_n) - f(w_n Z_n) \right| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n]
 \end{aligned}$$

$$\begin{aligned}
 & + \sup_{f \in \mathcal{F}} \int_{|R_n| > \epsilon} |f(w_n Z_n + R_n) - f(w_n Z_n)| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \\
 & \leq \sup_{f \in \mathcal{F}} \int_{|R_n| \leq \epsilon} \|f\|_{BL} |w_n Z_n + R_n - w_n Z_n| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \\
 & \quad + \sup_{f \in \mathcal{F}} \int_{|R_n| > \epsilon} (|f(w_n Z_n + R_n)| + |f(w_n Z_n)|) d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \\
 & \leq \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_n| \leq \epsilon} |R_n| d\mathbb{P}[\cdot | \mathcal{I}_n] + 2 \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_n| > \epsilon} d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \\
 & \leq \int_{|R_n| \leq \epsilon} \epsilon d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] + 2 \bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{I}_n] \leq \epsilon + 2 \bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{I}_n].
 \end{aligned}$$

In line with Xiong and Li (2008, Theorem 3.3), employing Markov's inequality we have

$$\bar{\mathbb{P}}[I \geq 2\epsilon] \leq \bar{\mathbb{P}}[\bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{I}_n] \geq \epsilon/2] \leq \frac{2}{\epsilon} \bar{\mathbb{P}}[|R_n| > \epsilon] \rightarrow 0$$

as $R_n = o_p(1)$ and hence $I = o_p(1)$. Consider II and let $K \geq 1$. We obtain

$$\bar{\mathbb{P}}[II \geq \epsilon] \leq \bar{\mathbb{P}}[||w_n|| \geq K] + \bar{\mathbb{P}}[II \geq \epsilon \cap ||w_n|| \leq K].$$

As $||w_n|| = O_p(1)$ the first term can be made arbitrarily small by choosing K large. For such K , consider the second term and note that

$$\begin{aligned}
 & \bar{\mathbb{P}}[II \geq \epsilon \cap ||w_n|| \leq K] \\
 & = \bar{\mathbb{P}}\left[\sup_{f \in \mathcal{F}} \left| \int [f(w_n Z_n) - f(w_n Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right| \geq \epsilon \cap ||w_n|| \leq K\right] \\
 & \leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_n) - g(Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right| \geq \epsilon \cap ||w_n|| \leq K\right] \\
 & \leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_n) - g(Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_n] \right| \geq \epsilon\right] \\
 & = \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int g d\left(\mathbb{P}[Z_n \leq \cdot | \mathcal{I}_n] - G_\infty\right) \right| \geq \epsilon\right]
 \end{aligned}$$

with $\mathcal{G} = \{g : \mathbb{R}^r \rightarrow \mathbb{R} \mid g(x) = f(w'x), \text{ for some } f \in \mathcal{F} \text{ and some } w \in \mathbb{R}^r \text{ with } ||w|| \leq$

$K\}$. We have that $\|\cdot\|_{BL}$ is uniformly bounded for \mathcal{G} since for every $g \in \mathcal{G}$

$$\begin{aligned} \|g\|_{BL} &= \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \frac{|w'x - w'y|}{\|x - y\|} \\ &\leq \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \|w\| \leq \|f\|_{BL} K \leq K. \end{aligned}$$

Thus, $\|g/K\|_{BL} \leq 1$ and it follows by $\bar{\mathbb{P}}[Z_n \leq \cdot | \mathcal{I}_n] \xrightarrow{P} G_\infty$ and Lemma 2.3 that

$$\begin{aligned} \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_n\| \leq K] &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int \frac{g}{K} d(\mathbb{P}[Z_n \leq \cdot | \mathcal{I}_n] - G_\infty) \right| \geq \frac{\epsilon}{K}\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{h \in \mathcal{H}} \left| \int h d(\mathbb{P}[Z_n \leq \cdot | \mathcal{I}_n] - G_\infty) \right| \geq \frac{\epsilon}{K}\right] \rightarrow 0, \end{aligned}$$

where \mathcal{H} is defined in Lemma 2.3. Thus, II is $o_p(1)$, which completes the proof. \square

Proof of Lemma 2.5. Take $\epsilon > 0$ and let F and G be cdfs on \mathbb{R} with $G(\tau - \epsilon) - \epsilon \leq F(\tau) \leq G(\tau + \epsilon) + \epsilon$ for all $\tau \in \mathbb{R}$. Fixing $u \in (\epsilon, 1 - \epsilon)$, we obtain

$$\inf \{ \tau \in \mathbb{R} : F(\tau) \geq u + \epsilon \} + \epsilon \tag{2.45}$$

$$\begin{aligned} &\geq \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) + \epsilon \geq u + \epsilon \} + \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} + \epsilon \\ &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau) \geq u \} \end{aligned} \tag{2.46}$$

$$\begin{aligned} &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) - \epsilon \geq u - \epsilon \} - \epsilon \\ &\geq \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u - \epsilon \} - \epsilon. \end{aligned} \tag{2.47}$$

Identifying (2.45), (2.46) and (2.47) as $F^{-1}(u + \epsilon) + \epsilon$, $G^{-1}(u)$ and $F^{-1}(u - \epsilon) - \epsilon$, respectively, completes the proof. \square

Proof of Lemma 2.6. Let $\epsilon, \eta > 0$. As G_n is stochastically uniformly equicontinuous, there exists a $\delta > 0$ and $\bar{n}_1 \in \mathbb{N}$ such that $\mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau - \tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon\right] < \eta$ for all $n \geq \bar{n}_1$. Take $\varkappa = \min(\delta/2, \epsilon)$. As $d_L(F_n, G_n) \xrightarrow{P} 0$ as

$n \rightarrow \infty$, there exists an \bar{n}_2 such that $\mathbb{P}[d_L(F_n, G_n) > \varkappa] < \eta$ for all $n \geq \bar{n}_2$. Let $\bar{n} = \max(\bar{n}_1, \bar{n}_2)$.

$$\begin{aligned}
 & \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} |F_n(\tau) - G_n(\tau)| > 2\epsilon\right] \\
 & \leq \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} |F_n(\tau) - G_n(\tau)| > 2\epsilon \cap d_L(F_n, G_n) \leq \varkappa\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
 & \leq \mathbb{P}\left[\varkappa + \sup_{\tau \in \mathbb{R}} |G_n(\tau \pm \varkappa) - G_n(\tau)| > 2\epsilon\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
 & \leq \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau - \tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] < 2\eta
 \end{aligned}$$

for all $n \geq \bar{n}$. As the choice of ϵ and η was arbitrary, the desired result follows. \square

Proof of Lemma 2.7. Since F_n and G_n merge in probability and $d_L \leq 2d_{BL}^{1/2}$ (cf. Huber, 2009, p. 36; Dudley, 2002, Theorem 11.3.3), we have $d_L(F_n, G_n) \xrightarrow{P} 0$. Let $u \in (0, 1)$ and take $\epsilon > 0$ sufficiently small satisfying $u \in (\epsilon, 1 - \epsilon)$. $\mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{I}_n]$ is $o_p(1)$ since for every $\delta > 0$ the Markov inequality implies $\mathbb{P}[\mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{I}_n] \geq \delta] \leq \frac{1}{\delta} \mathbb{P}[d_L(F_n, G_n) > \epsilon] \rightarrow 0$. Employing Lemma 2.5 we derive the following bounds:

$$\begin{aligned}
 & \mathbb{P}[M_n \leq G_n^{-1}(u) | \mathcal{I}_n] \\
 & \leq \mathbb{P}[M_n \leq G_n^{-1}(u) \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{I}_n] + \mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{I}_n] \\
 & \leq \mathbb{P}[M_n \leq F_n^{-1}(u + \epsilon) + \epsilon \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{I}_n] + o_p(1) \\
 & \leq \mathbb{P}[M_n \leq F_n^{-1}(u + \epsilon) + \epsilon | \mathcal{I}_n] + o_p(1) \\
 & = F_n(F_n^{-1}(u + \epsilon) + \epsilon) + o_p(1) = U_n
 \end{aligned}$$

$$\begin{aligned}
 & \mathbb{P}[M_n < G_n^{-1}(u) | \mathcal{I}_n] \\
 & \geq \mathbb{P}[M_n < G_n^{-1}(u) \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{I}_n] \\
 & \geq \mathbb{P}[M_n < F_n^{-1}(u - \epsilon) - \epsilon \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{I}_n] \\
 & \geq \mathbb{P}[M_n < F_n^{-1}(u - \epsilon) - \epsilon | \mathcal{I}_n] - \mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{I}_n] \\
 & = F_n(F_n^{-1}(u - \epsilon) - \epsilon) - o_p(1) = L_n,
 \end{aligned}$$

where $F_n(\cdot -)$ denotes the left limit of $F_n(\cdot)$. We show that L_n and U_n converge

in probability to u . Regarding the lower bound L_n we have

$$\begin{aligned}
 & \left| F_n(F_n^{-1}(u - \epsilon) - \epsilon -) - u \right| \\
 & \leq \left| F_n(F_n^{-1}(u - \epsilon) - \epsilon -) - F_n(F_n^{-1}(u - \epsilon) -) \right| \\
 & \quad + \left| F_n(F_n^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
 & \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau - \epsilon -) - F_n(\tau -) \right| + \left| F_n(F_n^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
 & \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau - \epsilon -) - F_n(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| F_n(\tau) - F_n(\tau -) \right| + \epsilon \\
 & \leq 4d_K(F_n, G_n) + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau - \epsilon -) - G_n(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau) - G_n(\tau -) \right| + \epsilon,
 \end{aligned}$$

where the third inequality is due to Cavaliere, Georgiev, and Taylor (2013, p. 217). As $d_L(F_n, G_n) \xrightarrow{p} 0$ and G_n is stochastically uniformly equicontinuous, Lemma 2.6 implies that $d_K(F_n, G_n) \xrightarrow{p} 0$. Further, $\sup_{\tau \in \mathbb{R}} |G_n(\tau - \epsilon -) - G_n(\tau -)| = o_p(1)$ and $\sup_{\tau \in \mathbb{R}} |G_n(\tau) - G_n(\tau -)| = o_p(1)$ by stochastic uniform equicontinuity completing $L_n \xrightarrow{p} u$. Regarding the upper bound U_n we have

$$\begin{aligned}
 & \left| F_n(F_n^{-1}(u + \epsilon) + \epsilon) - u \right| \\
 & \leq \left| F_n(F_n^{-1}(u + \epsilon) + \epsilon) - F_n(F_n^{-1}(u + \epsilon)) \right| \\
 & \quad + \left| F_n(F_n^{-1}(u + \epsilon)) - F_n(F_n^{-1}(u + \epsilon) -) \right| \\
 & \quad + \left| F_n(F_n^{-1}(u + \epsilon) -) - (u + \epsilon) \right| + \epsilon \\
 & \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau + \epsilon) - F_n(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| F_n(\tau) - F_n(\tau -) \right| + \epsilon \\
 & \leq 6d_K(F_n, G_n) + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau + \epsilon) - G_n(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| G_n(\tau) - G_n(\tau -) \right| + \epsilon,
 \end{aligned}$$

where all terms on the right hand side are $o_p(1)$ such that $U_n \xrightarrow{p} u$. We obtain

$$\underbrace{L_n}_{\xrightarrow{p} u} \leq \mathbb{P}[M_n < G_n^{-1}(u) | \mathcal{I}_n] \leq \mathbb{P}[M_n \leq G_n^{-1}(u) | \mathcal{I}_n] \leq \underbrace{U_n}_{\xrightarrow{p} u},$$

which implies that $\mathbb{P}[M_n < G_n^{-1}(u) | \mathcal{I}_n]$ and $\mathbb{P}[M_n \leq G_n^{-1}(u) | \mathcal{I}_n]$ converge in probability to u for arbitrary $u \in (0, 1)$; in particular γ_1 and $1 - \gamma_2$. It follows that

$$\mathbb{P}[G_n^{-1}(\gamma_1) \leq M_n \leq G_n^{-1}(1 - \gamma_2) | \mathcal{I}_n]$$

$$= \mathbb{P}[M_n \leq G_n^{-1}(1 - \gamma_2) | \mathcal{I}_n] - \mathbb{P}[M_n < G_n^{-1}(\gamma_1) | \mathcal{I}_n] \xrightarrow{P} 1 - \gamma_2 - \gamma_1. \quad \square$$

Proof of Lemma 2.8. Let $\epsilon, \eta > 0$ and set $\bar{\epsilon} = \min\{\epsilon, u, 1 - u\}/2$. Since G_n^{-1} is pointwise equicontinuous at u , there exist a $\delta > 0$ and an \bar{n}_1 such that $\mathbb{P}[\sup_{|u-v|<\delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \bar{\epsilon}] < \eta$ for all $n \geq \bar{n}_1$. Take $\varkappa = \min\{\delta/2, \bar{\epsilon}\}$. As $d_L(F_n, G_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$, there exists an \bar{n}_2 such that $\mathbb{P}[d_L(F_n, G_n) > \varkappa] < \eta$ for all $n \geq \bar{n}_2$. We obtain

$$\begin{aligned} & \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\epsilon\right] \leq \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\bar{\epsilon}\right] \\ & \leq \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\bar{\epsilon} \cap d_L(F_n, G_n) \leq \varkappa\right] + \mathbb{P}\left[d_L(F_n, G_n) > \varkappa\right] \\ & \leq \mathbb{P}\left[\varkappa + |G_n^{-1}(u \pm \varkappa) - G_n^{-1}(u)| > 2\bar{\epsilon}\right] + \mathbb{P}\left[d_L(F_n, G_n) > \varkappa\right] \\ & \leq \mathbb{P}\left[\sup_{|u-v|<\delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \bar{\epsilon}\right] + \mathbb{P}\left[d_L(F_n, G_n) > \varkappa\right] < 2\eta \end{aligned}$$

for all $n \geq \bar{n} = \max(\bar{n}_1, \bar{n}_2)$, where the third inequality follows from Lemma 5 and $u \in (\bar{\epsilon}, 1 - \bar{\epsilon}) \subseteq (\varkappa, 1 - \varkappa)$. As ϵ and η were arbitrarily chosen, this completes the proof. \square

2.B.2 Proofs of Corollaries

Proof of Corollary 2.1. Statement 1.(i) follows immediately from Theorem 2.2 (statement 1.(i)) and $\widehat{F_n^{2IP}}(\cdot)$ equals to $\Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$. Regarding claim 1.(ii), it is sufficient to show that $1/\hat{v}_n^{2IP} = O_p(1)$ implies that $\Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$ is stochastically uniformly equicontinuous by Theorem 2.2 (statement 1.(ii)). Since $1/\hat{v}_n^{2IP} = O_p(1)$ by assumption, we have for all $\varkappa > 0$, there exist $K = K(\varkappa)$ and $\bar{n} = \bar{n}(\varkappa)$ such that $\mathbb{P}[1/\hat{v}_n^{2IP} > K] < \varkappa$ for all $n > \bar{n}$. Let ϕ denote the standard normal density. Taking $\delta = \frac{\epsilon}{\phi(0)\sqrt{K}}$, we obtain

$$\begin{aligned} & \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau': |\tau - \tau'| < \delta} |\Phi(\tau' / \sqrt{\hat{v}_n^{2IP}}) - \Phi(\tau / \sqrt{\hat{v}_n^{2IP}})| > \epsilon\right] \\ & \leq \mathbb{P}\left[\phi(0)\delta / \sqrt{\hat{v}_n^{2IP}} > \epsilon\right] = \mathbb{P}\left[1/\hat{v}_n^{2IP} > K\right] < \varkappa \end{aligned}$$

for all $n > \bar{n}$ such that the stochastic uniform equicontinuity condition holds.

Statement 2.(i) follows from Theorem 2.2 (statement 2.(i)) and $\widehat{F_n^{SPL}}(\cdot)$ equals

$\Phi(\cdot / \sqrt{\hat{v}_n^{SPL}})$. Claim 2.(ii) is proven analogously to the claim of 1.(ii) replacing \hat{v}_n^{2IP} with \hat{v}_n^{SPL} . \square

Proof of Corollary 2.2. In the proof of Theorem 2.3 it is shown that $\widehat{F}_n^{SPL}(\cdot)$ and $\widehat{F}_n^{STA}(\cdot)$ merge in probability, which simplify to $\Phi(\cdot / \sqrt{\hat{v}_n^{SPL}})$ and $\Phi(\cdot / \sqrt{\hat{v}_n^{STA}})$, respectively, under Assumption 2.5. It remains to show that the quantile function $\widehat{F}_n^{SPL^{-1}}(u) = \sqrt{\hat{v}_n^{SPL}}\Phi^{-1}(u)$ is stochastically pointwise equicontinuous at $u = \gamma_1, 1 - \gamma_2$. First, we show that $\hat{v}_n^{SPL} = O_p(1)$. The triangle inequality implies $\hat{v}_n^{SPL} \leq v_n^{SPL} + |\hat{v}_n^{SPL} - v_n^{SPL}|$, where $v_n^{SPL} = \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta'} \Upsilon_0 \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta}$ is $O_p(1)$ by Assumptions 2.1.(iii) and 2.1.(v). Further, for an arbitrary $\varepsilon > 0$, we get

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E}))}{\partial \theta} \right\| \geq \varepsilon \right] \\ & \leq \mathbb{P} \left[\left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \tilde{\theta}_n)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0 \right\| \geq \varepsilon \cap \tilde{\theta}_n \in \mathcal{V}(\theta_0) \right] \\ & \quad + \mathbb{P}[\tilde{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0 \right\| \geq \varepsilon \right] + \mathbb{P}[\tilde{\theta}_n \notin \mathcal{V}(\theta_0)], \end{aligned}$$

where $\tilde{\theta}_n$ lies between $\hat{\theta}(\mathbf{X}_{1:n_E})$ and θ_0 . The first term of the sum vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ by Assumptions 2.1.(iv) and 2.1.(v) and $\left\| \hat{\theta}(\mathbf{X}_{1:n_E}) - \theta_0 \right\| = O_p(m_n^{-1})$ by Assumptions 2.1.(i) and 2.2.(i). Further, since $\hat{\theta}(\mathbf{X}_{1:n_E}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\tilde{\theta}_n \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n+1}^s(\mathbf{X}_{n_P:n}^c; \hat{\theta}(\mathbf{X}_{1:n_E}))}{\partial \theta} \right\| = o_p(1)$ follows. Together with $\hat{\Upsilon}(\mathbf{X}_{1:n_E}) \xrightarrow{p} \Upsilon_0$, it implies $|\hat{v}_n^{SPL} - v_n^{SPL}| = o_p(1)$ and hence $\hat{v}_n^{SPL} = O_p(1)$. Next, we show that the stochastic pointwise equicontinuity condition is satisfied. For $K > 0$, we get

$$\begin{aligned} & \mathbb{P} \left[\sqrt{\hat{v}_n^{SPL}} \sup_{v: |u-v| < \delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \varepsilon \right] \\ & \leq \mathbb{P} \left[\sqrt{K} \sup_{v: |u-v| < \delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \varepsilon \right] + \mathbb{P}[\hat{v}_n^{SPL} > K]. \end{aligned}$$

K can be chosen such that the last term is arbitrary small for large n as $\hat{v}_n^{SPL} = O_p(1)$. Given K , the first term is 0 by the choice of δ and continuity of Φ^{-1} . \square

Chapter 3

A General Framework for Prediction in Time Series Models

In this chapter we propose a general framework to analyze prediction in time series models and show how a wide class of popular time series models satisfies this framework. Our framework coincides with that of Chapter 2 where we establish the validity of conditional confidence intervals for predictions made in this framework. In the chapter at hand we formally verify the high-level assumptions of the previous chapter for the aforementioned time series models. The current chapter therefore complements the results of Chapter 2 by providing practically relevant applications of its theory.¹

¹This chapter is based on the paper Beutner, Heinemann, and Smeekes (2019a).

3.1 Introduction

In time series prediction one is frequently interested in objects that do not only depend on parameters but also on the time series' past. Popular examples are conditional means or conditional variances. Analyzing predictions in this context involves a fundamental issue that is well-recognized in the econometric literature. It stems from the fact that on the one hand one must *condition* on the sample as the past informs about the present and future, yet on the other hand one must treat the data up to now as random to take into account parameter uncertainty. Nevertheless the issue is often ignored in standard practice or bypassed by assuming two independent processes with the same stochastic structure, using one for the conditioning and one for the estimation of the parameters (see Section 2.2). While the latter is a mathematically convenient assumption, it is rarely satisfied in practice. An alternative, more realistic approach is based on sample-splitting, in which one splits the sample into two (asymptotically) independent subsamples.

In this chapter we provide a general framework to analyze prediction in time series models. We revisit the set of high-level assumptions of Chapter 2 under which we establish the validity of conditional confidence intervals for predictions while demonstrating an asymptotic equivalence of two-independent processes and the sample-split approach. We show how a wide class of popular time series models satisfies this framework. In particular, we consider autoregressive moving-average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) type models and formally verify the high-level assumptions of the previous chapter. Therefore the current paper complements the results of Chapter 2 by providing practically relevant applicants to its theory.

The rest of this chapter is organized as follows. The general framework to analyze prediction in time series models is proposed in Section 3.2. In Sections 3.3 and 3.4 we revisit the leading examples of Chapter 2, i.e. the simple case of a conditional mean in an AR(1) and the conditional variance in a GARCH(1,1) model. In Section 3.5 we focus on the conditional mean in a slightly more general model: the ARMA(1,1) with drift. Section 3.6 studies the conditional volatility in a threshold GARCH model. Concluding remarks are presented in Section 3.7.

3.2 General Framework

Let $\{X_t\}$ be a univariate stochastic process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote the relevant parameter (vector) by θ_0 , which belongs to some set $\Theta \subseteq \mathbb{R}^r$, $r \in \mathbb{N}$. The general framework involves inference on objects, which are a function not only of the parameter but also the time series' past. Mathematically, such object can be written as follows:

$$\psi_{n+1} = \psi(X_n, X_{n-1}, \dots; \theta_0) \quad (3.1)$$

for some function $\psi : \mathbb{R}^\infty \times \Theta \rightarrow \mathbb{R}$. Such prediction function can generally not be determined completely given a sample X_1, \dots, X_n . Replacing the unknown presample values by arbitrary starting values $\{s_t\}$, yields the following approximation:

$$\psi_{n+1}^s(\mathbf{X}_{1:n}; \theta_0) = \psi(X_n, X_{n-1}, \dots, X_1, s_0, s_{-1}, \dots; \theta_0), \quad (3.2)$$

where $\mathbf{X}_{t_1:t_2} = (X_{t_1}, \dots, X_{t_2})'$ for any integers $1 \leq t_1 \leq t_2 \leq n$. To estimate the prediction function in practice, the standard approach is to replace the unknown parameter θ_0 by an estimator $\hat{\theta}(\mathbf{X}_{1:n})$. Conditioning on the entire sample for the evaluation of the prediction function entails that there is no randomness to account for parameter uncertainty, which highlights the severity of the fundamental issue at hand. The issue is frequently bypassed by making the unrealistic assumption of observing two independent processes, where one is used for the evaluation of the prediction function and the other for parameter estimation.

An alternative, more realistic approach is based on splitting the sample into two (asymptotically) independent subsamples. The successive decline of the influence of past observations, which motivated the approximation in (3.2), entails that

$$\psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) = \psi(X_n, X_{n-1}, \dots, X_{t_1}, c_{t_1-1}, \dots, c_1, s_0, s_{-1}, \dots; \theta_0) \quad (3.3)$$

serves as an approximation for (3.2) (and hence for (3.1)) for an appropriate choice of t_1 . Here $\mathbf{X}_{t_1:n}^c = (c_1, \dots, c_{t_1-1}, X_{t_1}, \dots, X_n)'$ is a vector where a subsample is substituted by a sequence of constants $\{c_t\}$, in a similar way as done for the starting values. Denoting the appropriate choice of t_1 by n_P , which indicates the starting point of the prediction sample, the sample-split estimator is obtained by replacing θ_0 in (3.3) by an estimator $\hat{\theta}(\mathbf{X}_{1:n_E})$, where n_E stands for the end of the

estimation sample. Choosing n_E to satisfy $1 < n_E < n_P \leq n$ yields an estimation subsample that does not overlap with the subsample used for prediction.

In the following sections we revisit the set of high-level assumptions of the previous chapter (Assumptions 2.1 to 2.5) under which we establish the validity of conditional confidence intervals for predictions while demonstrating an asymptotic equivalence of two-independent processes and the sample-split approach. In particular we formally verify the high-level assumptions of Chapter 2 for a wide class of popular time series models satisfying the general framework. Since the subsequently considered ARMA and GARCH models exhibit an exponential decay in memory we henceforth set $l_n = \log n$ with regard to Assumption 2.1. Further, we constrain ourselves to \sqrt{n} -consistent estimators of the parameters such that $m_n = \sqrt{n}$ throughout the chapter.

3.3 Conditional Mean in an AR(1)

3.3.1 Model Description

An autoregressive model represents a process in terms of its lagged value(s) and some stochastic innovation process. The first order autoregressive process without drift is defined by the following recursion

$$X_t = \beta_0 X_{t-1} + \varepsilon_t, \quad (3.4)$$

for $t \in \mathbb{Z}$, where the parameter $\beta_0 \in \Theta$ satisfies $|\beta_0| < 1$ and $\{\varepsilon_t\}$ is a sequence of innovations. Subsequently, we make the following assumptions.

Assumption 3.1. (*AR(1)-model*)

- (i) (*Compactness*) Θ is compact;
- (ii) (*Interior*) $\beta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ ;
- (iii) (*Causality*) $|\beta| < 1$ for all $\beta \in \Theta$;
- (iv) (*Innovations*) ε_t are i.i.d. from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive on \mathbb{R} ;

Θ is assumed to be compact in Assumption 3.1.(i), which holds true, for instance, if it is of the form $\Theta = \{\beta' \in \mathbb{R} : |\beta| \leq 1 - \delta\}$, where $\delta > 0$ is a sufficiently small constant. Assumption 3.1.(ii) states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The causality condition is stated in 3.1.(iii). Assumption 3.1.(iv) imposes further restrictions on the distribution of the innovation process. Next, we turn to the estimation of the model.

3.3.2 Estimation

To estimate the model in equation (3.4), we employ the OLS estimator given by

$$\hat{\beta}(\mathbf{X}_{1:n}) = \sum_{t=2}^n X_t X_{t-1} / \sum_{t=2}^n X_{t-1}^2 \quad (3.5)$$

As the sample size grows large, the OLS estimator approaches a normal distribution under regulatory conditions.

Theorem 3.1. (*Hamilton, 1994*) *Under Assumption 3.1*

$$\sqrt{n}(\hat{\beta}(\mathbf{X}_{1:n}) - \beta_0) \xrightarrow{d} N(0, \sigma_\beta^2) \quad (3.6)$$

with $\sigma_\beta^2 = 1 - \beta_0^2$.

3.3.3 Mapping

The mapping of the AR(1) process into the general setup is straightforward: β_0 corresponds to θ_0 and the conditional mean of X_{n+1} is equal to

$$\psi_{n+1} = \psi(X_n, X_{n-1}, \dots; \theta_0) = \beta_0 X_n. \quad (3.7)$$

3.3.4 Verification of Assumptions

Assumption 2.1

For Assumption 2.1.(i) to be met, we consider the OLS estimator in (3.5), whose asymptotic distribution is specified in Theorem 3.1.

As the function $\psi(\dots; \theta)$ given in (3.7) is continuous on Θ and twice differentiable on $\overset{\circ}{\Theta}$, Assumption 2.1.(ii) is met.

Consider Assumption 2.1.(iii) and notice that the gradient simplifies to

$$\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} = X_n.$$

Clearly, X_n is $O_p(1)$ since the process $\{X_t\}$ is strictly stationary; see also Assumption 2.3.(iii), which is verified below.

The condition in Assumption 2.1.(iv) is met as

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = 0.$$

Regarding Assumption 2.1.(v), we obtain for $t_1 < n$

$$m_n \left(\psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) - \psi(X_n, X_{n-1}, \dots; \theta_0) \right) = \sqrt{n}(\beta_0 X_n - \beta_0 X_n) = 0$$

and

$$\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| = |X_n - X_n| = 0$$

as well as

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = |0 - 0| = 0,$$

which completes the verification of Assumption 2.1.

Assumption 2.3

The condition in Assumption 2.3.(i) is satisfied for instance by $n_E(n) \sim n - \lfloor n^b \rfloor$ and $n_P(n) \sim n - \lfloor n^a \rfloor$ with $0 < a < b < 1$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x .

The process $\{X_t\}$ is strictly stationary since $|\beta_0| < 1$ and $\mathbb{E} \log^+ |\varepsilon_t| \leq \mathbb{E} |\varepsilon_t| < \infty$, where $\log^+ x = \max\{\log x, 0\}$ (Bougerol and Picard, 1992, Theorem 4.1).

The process $\{X_t\}$ is β -mixing with exponential decay (Mokkadem, 1988, Theorem 1'). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 2.3.(iii) is met with regard to Remark 2.3 and noting that $n_P(n) - n_E(n) \sim \lfloor n^b \rfloor - \lfloor n^a \rfloor \rightarrow \infty$ as $n \rightarrow \infty$. For alternative mixing results we refer to Davidson (1994, Theorem 14.9) or Andrews (1983, Theorem 1).

Assumptions 2.4 and 2.5

Assumption 2.4 is implied by Assumption 2.5, which, in turn, is verified by Theorem 3.1 and $\hat{\sigma}_\beta^2(\mathbf{X}_{1:n}) = 1 - \hat{\beta}(\mathbf{X}_{1:n})^2 \xrightarrow{P} \sigma_\beta^2$.

Assumptions within Corollary 2.1

We show $1/\hat{v}_n^{2IP} = O_p(1)$. By independence of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, the distribution of $X_n = \sum_{k=0}^{\infty} \beta_0^k \varepsilon_{n-k}$ is equal to $\mathbb{P}[X_n \leq \cdot] = \mathbb{P}[\varepsilon_n \leq \cdot] \times \mathbb{P}[\beta_0 \varepsilon_{n-1} \leq \cdot] \times \mathbb{P}[\beta_0^2 \varepsilon_{n-2} \leq \cdot] \times \dots$. As $\mathbb{P}[\varepsilon_t \leq \tau]$ is continuous and non-degenerate, so is $\mathbb{P}[X_n \leq \tau]$, which does not depend of n as $\{X_t\}_{t \in \mathbb{Z}}$ is strictly stationary. It follows that X_n is bounded away from zero. Further, write $\hat{v}_n^{2IP} = X_n^2 \hat{\sigma}_\beta^2(\mathbf{X}_{1:n}) = X_n^2 \sigma_\beta^2 + S_n$ and note that $S_n = X_n^2 (\hat{\sigma}_\beta^2(\mathbf{X}_{1:n}) - \sigma_\beta^2) = o_p(1)$. For every $\epsilon > 0$, we have

$$\begin{aligned} \mathbb{P}[\hat{v}_n^{2IP} \geq \epsilon] &\geq \mathbb{P}[X_n^2 \sigma_\beta^2 + S_n \geq \epsilon \cap |S_n| \leq \epsilon] \\ &\geq \mathbb{P}[X_n^2 \sigma_\beta^2 \geq 2\epsilon \cap |S_n| \leq \epsilon] \\ &\geq \mathbb{P}[X_n^2 \sigma_\beta^2 \geq 2\epsilon] - \mathbb{P}[|S_n| > \epsilon], \end{aligned}$$

where the last inequality follows from $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] - \mathbb{P}[B^c]$. Fix $\delta > 0$; since X_n and hence X_n^2 are bounded away from zero and $\sigma_\beta^2 > 0$, there exists an $\epsilon = \epsilon(\delta)$ such that $\mathbb{P}[X_n^2 \geq 2\epsilon/\sigma_\beta^2] \geq 1 - \delta/2$. For such ϵ , there exists an $\bar{n} = \bar{n}(\epsilon(\delta), \delta) = \bar{n}(\delta)$ such that $\mathbb{P}[|S_n| > \epsilon] < \delta/2$ for all $n \geq \bar{n}$ since $S_n = o_p(1)$. It follows that $\mathbb{P}[\hat{v}_n^{2IP} \geq \epsilon] \geq 1 - \delta$ for all $n \geq \bar{n}$. As $\delta > 0$ was arbitrarily chosen, this completes the proof of \hat{v}_n^{2IP} being bounded away from zero. The proof of \hat{v}_n^{SPL} being bounded away from zero is analogous and hence omitted.

3.4 Conditional Variance in a GARCH(1,1)

3.4.1 Model Description

Autoregressive conditional heteroscedasticity models were originally introduced by Engle (1982) and extended to GARCH models by Bollerslev (1986). The model reflects the predominant characteristics of financial returns justifying its popularity among practitioners. The model's temporal dependence structure captures the slow decaying autocorrelations of absolute financial returns, also known as volatility clustering. The GARCH(1, 1) process $\{X_t\}$ is defined by

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \end{aligned} \tag{3.8}$$

for all $t \in \mathbb{Z}$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ are non-negative parameters in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. In the traditional GARCH model, Bollerslev (1986) assumed the innovations $\{\varepsilon_t\}$ to be independent following a standard normal distribution. The normality assumption is commonly relaxed to account for stylized statistical properties of financial returns such as *skewness* due to leverage effects and *kurtosis*, also known as fat tails. We denote by $\theta = (\omega, \alpha, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 3.2. (*GARCH(1,1)-model*)

- (i) (*Compactness*) Θ is compact;
- (ii) (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;
- (iii) (*Non-negativity*) $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ for all $\theta \in \Theta$;
- (iv) (*Strict stationarity*) $\mathbb{E}[\ln(\alpha_0 \varepsilon_t^2 + \beta_0)] < 0$ and $\beta < 1$ for all $\theta \in \Theta$;
- (v) (*Roots*) $\alpha_0 z > 0$ and $1 - \beta_0 z > 0$ have no common root, and $\alpha_0 > 0$;
- (vi) (*Innovations*) ε_t are i.i.d. from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = 1$ and $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive in a neighborhood of zero;

Θ is assumed to be compact in Assumption 3.2.(i), which holds true, for instance, if it is of the form $\Theta = [\delta, 1/\delta] \times [0, 1/\delta] \times [0, 1 - \delta]$, where $\delta \in (0, 1)$ is a

sufficiently small constant. Assumption 3.2.(ii) states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The non-negativity constraints in 3.2.(iii) are standard ensuring the conditional variance to be strictly positive. Assumption 3.2.(iv) is necessary and sufficient for $\{X_t\}$ being strictly stationary (cf. Francq and Zakoïan, 2011, Theorem 2.1). The root condition in 3.2.(v) guarantees that the GARCH model is irreducible. Assumption 3.2.(vi) imposes further restrictions on the moments and density of the innovation process. Next, we turn to the estimation of the model in (3.8).

3.4.2 Estimation

We consider the quasi maximum likelihood (QML) estimator proposed by Francq and Zakoïan (2004) to estimate the GARCH(1,1) model. For a generic $\theta \in \Theta$ we set

$$\sigma_{t+1}^2(\theta) = \sum_{k=0}^{\infty} \beta^k (\omega + \alpha X_{t-k}^2) \quad (3.9)$$

and note that $\sigma_{t+1}^2 = \sigma_{t+1}^2(\theta_0)$. Replacing the unknown presample observations by arbitrary values, say s_t , $t \leq 0$, we denote the modified version of (3.9) by $\tilde{\sigma}_{t+1}^2(\theta)$. Then the QML estimator of θ_0 is defined as any measurable solution $\hat{\theta}(\mathbf{X}_{1:n})$ of

$$\hat{\theta}(\mathbf{X}_{1:n}) = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta; \mathbf{X}_{1:n}) \quad (3.10)$$

with

$$\tilde{L}_n(\theta; \mathbf{X}_{1:n}) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\theta)}} \exp\left(-\frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)}\right).$$

Assumption 3.2 implies that the estimator follows asymptotically a normal distribution.

Theorem 3.2. (*Francq and Zakoïan, 2004*) *Under Assumption 3.2*

$$\sqrt{n}(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0), \quad (3.11)$$

where $\Upsilon_0 = (\mathbb{E}[\varepsilon_t^4] - 1) \mathbb{E} \left[\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} \right]^{-1}$ and $\sigma_t^2(\theta)$ is given in (3.9).

It is worth stressing that Υ_0 does not only depend on θ_0 but also on some nuisance parameters such as $\mathbb{E}[\varepsilon_t^4]$.

3.4.3 Mapping

Having described the model and its estimation, we turn to map the model into the general setup of Chapter 2. The conditional variance σ_{n+1}^2 is equal to

$$\psi_{n+1} = \psi(X_n, X_{n-1}, \dots; \theta_0) = \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0 X_{n-k}^2). \quad (3.12)$$

To verify Assumption 2.1 the first and second derivatives of $\psi(X_n, X_{n-1}, \dots; \theta)$ w.r.t. θ are needed. The first order derivatives are

$$\begin{aligned} \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega} &= \frac{1}{1 - \beta}, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha} &= \sum_{k=0}^{\infty} \beta^k X_{n-k}^2, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} (\omega + \alpha X_{n-k}^2), \end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned} \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} &= 0, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha} &= 0, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} &= 0, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} &= -\frac{1}{(1 - \beta)^2}, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{n-k}^2, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} &= \sum_{k=2}^{\infty} k(k-1) \beta^{k-2} (\omega + \alpha X_{n-k}^2). \end{aligned}$$

3.4.4 Verification of Assumptions

Before turning to the verification of the assumptions of Chapter 2, note that the strict stationarity condition implies the existence of fractional moments: there exists an $s \in (0, 1)$ such that $\mathbb{E}X_t^{2s} < \infty$ (Nelson, 1990, Theorem 2). For such $s \in (0, 1)$ the following elementary inequalities hold: $(a + b)^s \leq a^s + b^s$ for all $a, b \geq 0$ and $c^s \leq c$ for all $c \geq 1$.

Assumption 2.1

For Assumption 2.1.(i) to be met, we consider the QML estimator of Francq and Zakoïan (2004), whose asymptotic distribution is specified in Theorem 3.2.

As the function $\psi(\dots; \theta)$, given in (3.12), is continuous on Θ and twice differentiable on $\mathring{\Theta}$, Assumption 2.1.(ii) is satisfied.

Consider Assumption 2.1.(iii) and note that

$$\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} = \frac{1}{1 - \beta_0}$$

is trivially $O(1)$. For showing $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha}\right| \geq M\right] < \epsilon$ for n sufficiently large. Employing the Markov inequality, we obtain

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{n-k}^2\right)^s\right] \\ &\leq \frac{1}{M^s} \sum_{k=0}^{\infty} \beta_0^{sk} \mathbb{E}X_t^{2s} = \frac{\mathbb{E}X_t^{2s}}{(1 - \beta_0^s)M^s} \end{aligned}$$

such that $M > \left(\frac{\mathbb{E}X_t^{2s}}{(1 - \beta_0^s)\epsilon}\right)^{1/s}$ gives the desired result. Similarly, we get

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=1}^{\infty} k \beta_0^{k-1} (\omega_0 + \alpha_0 X_{n-k}^2)\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\sum_{k=1}^{\infty} k \beta_0^{s(k-1)} (\omega_0^s + \alpha_0^s X_{n-k}^{2s})\right] = \frac{\omega_0^s + \alpha_0^s \mathbb{E}X_t^{2s}}{M^s(1 - \beta_0^s)^2} \end{aligned}$$

such that $M > \left(\frac{\omega_0^s + \alpha_0^s \mathbb{E} X_t^{2s}}{\epsilon(1-\beta_0^s)^2} \right)^{1/s}$ establishes $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} = O_p(1)$, which completes the verification of Assumption 2.1.(iii).

Focusing on Assumption 2.1.(iv) we notice that

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} \right| = \frac{1}{(1 - \beta_{\sup})^2} = O(1),$$

where $\beta_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \beta$. To show $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| = O_p(1)$, we need to find an M for every $\epsilon > 0$ such that $\mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] < \epsilon$ holds. We find

$$\begin{aligned} \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^2 \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\sup}^{s(k-1)} X_{n-k}^{2s} \right] \leq \frac{1}{M^s} \sum_{k=1}^{\infty} k \beta_{\sup}^{s(k-1)} \mathbb{E} X_t^{2s} = \frac{\mathbb{E} X_t^{2s}}{M^s (1 - \beta_{\sup}^s)^2}. \end{aligned}$$

Taking $M > \left(\frac{\mathbb{E}X_t^{2s}}{\epsilon(1-\beta_{\sup}^s)^2} \right)^{1/s}$ leads to the desired result. Similarly, we have

$$\begin{aligned} & \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} \right| \geq M \right] \\ &= \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=2}^{\infty} k(k-1) \beta_{\sup}^{k-2} (\omega_{\sup} + \alpha_{\sup} X_{n-k}^2) \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=2}^{\infty} k^s (k-1)^s \beta_{\sup}^{s(k-2)} (\omega_{\sup}^s + \alpha_{\sup}^s X_{n-k}^{2s}) \right] \\ &\leq \frac{1}{M^s} \sum_{k=2}^{\infty} k(k-1) \beta_{\sup}^{s(k-2)} (\omega_{\sup}^s + \alpha_{\sup}^s \mathbb{E}X_t^{2s}) = \frac{2(\omega_{\sup}^s + \alpha_{\sup}^s \mathbb{E}X_t^{2s})}{M^s(1-\beta_{\sup}^s)^3}, \end{aligned}$$

where $\omega_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \omega$ and $\alpha_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha$. Taking M larger than $\left(\frac{2(\omega_{\sup}^s + \alpha_{\sup}^s \mathbb{E}X_t^{2s})}{\epsilon(1-\beta_{\sup}^s)^3} \right)^{1/s}$ completes the verification of Assumption 2.1.(iv).

Regarding Assumption 2.1.(v) we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e. $c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_n, X_{n-1}, \dots; \theta) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta) = \sum_{k=n-t_1+1}^{\infty} \beta^k \alpha X_{n-k}^2.$$

We have

$$\begin{aligned} & m_n \left(\psi(X_n, X_{n-1}, \dots; \theta_0) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) \right) \\ &= \sqrt{n} \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^k \alpha_0 X_{t_1-k}^2. \end{aligned} \tag{3.13}$$

Clearly, the sum is of order $O_p(1)$. Further, for any $t_1 \geq 1$ such that $(n-t_1)/l_n \rightarrow \infty$ we get $\sqrt{n} \beta_0^{n-t_1} \rightarrow 0$. Hence, (3.13) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

and

$$\left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \alpha} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha} \right| = \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^2$$

being $o_p(1)$ since the sum is $O_p(1)$ and $\beta_0^{n-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} \right| = \sum_{k=n-t_1+1}^{\infty} k \beta_0^{k-1} \alpha_0 X_{n-k}^2 \\ & = (n-t_1) \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^{k-1} \alpha_0 X_{t_1-k}^2 + \beta_0^{n-t_1} \sum_{k=1}^{\infty} k \beta_0^{k-1} \alpha_0 X_{t_1-k}^2 \end{aligned}$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| = (0, 0, 0), \\ & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^2} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} \right| = 0 \end{aligned}$$

and

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha \partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| = \sum_{k=n-t_1+1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^2 \\ & = (n-t_1) \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} \beta_{\sup}^{k-1} X_{t_1-k}^2 + \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{t_1-k}^2 \end{aligned}$$

is $o_p(1)$ by previous arguments noting that $\beta_{\sup} \in (0, 1)$. Similarly, it can be shown that

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \beta \partial \beta'} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta \partial \beta'} \right| \\ & = \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=n-t_1+1}^{\infty} k(k-1) \beta^{k-2} \alpha X_{n-k}^2 \right| \\ & \leq \sum_{k=n-t_1+1}^{n-1} k(k-1) \beta_{\sup}^{k-2} \alpha_{\sup} X_{n-k}^2 \end{aligned}$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 2.3

The condition in Assumption 2.3.(i) is satisfied for instance by $n_E(n) \sim n - \lfloor n^b \rfloor$ and $n_P(n) \sim n - \lfloor n^a \rfloor$ with $0 < a < b < 1$.

With regard to Assumption 3.2.(iv), $\{X_t\}$ is a strictly stationary process such that Assumption 2.3.(ii) is satisfied.

The process $\{X_t\}$ is β -mixing with exponential decay (Francq and Zakoïan, 2011, Theorem 3.4). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 2.3.(iii) is met with regard to Remark 2.3 noting that $n_P(n) - n_E(n) \rightarrow \infty$.

Assumptions 2.4 and 2.5

Assumption 2.4 is implied by Assumption 2.5, which, in turn, is verified by Theorem 3.2 and the consistent² estimator

$$\begin{aligned} \hat{\Upsilon}(\mathbf{X}_{1:n}) &= \left(\frac{1}{n} \sum_{t=1}^n \frac{X_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:n}))} - 1 \right) \\ &\quad \times \left(\frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:n}))} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta'} \right)^{-1}. \end{aligned}$$

Assumptions within Corollary 2.1

To show $1/\hat{v}_n^{2IP} = O_p(1)$, recall that $\hat{v}_n^{2IP} = v_n^{2IP} + o_p(1)$ (see proof of Corollary 2.2) and define $\kappa = \text{eig}_{\min} \Upsilon_0$, the minimum eigenvalue of Υ_0 . Since Υ_0 is positive definite, we have $\kappa > 0$ such that $1/\hat{v}_n^{2IP} = O_p(1)$ is implied by

$$\hat{v}_n^{2IP} + o_p(1) = v_n^{2IP} \geq \kappa \left\| \frac{\partial \psi_{n+1}}{\partial \theta} \right\|^2 \geq \kappa \left\| \frac{\partial \psi_{n+1}}{\partial \omega} \right\|^2 \geq \kappa.$$

²A formal proof of consistency under Assumption 3.2 is along the lines of the intermediary results (iv) and (vi) included in Francq and Zakoïan (2004, Theorem 2.2).

Similarly, we obtain $\hat{v}_n^{SPL} + o_p(1) \geq \kappa$ such that $1/\hat{v}_n^{SPL} = O_p(1)$.

3.5 Conditional Mean in an ARMA(1,1)

3.5.1 Model Description

The ARMA model was popularized by the classical book of Box and Jenkins (1971). It represents a stationary stochastic process in terms of an autoregressive and a moving-average part. The ARMA(1, 1) process with drift is given by

$$X_t - \omega_0 = \alpha_0 \varepsilon_{t-1} + \beta_0 (X_{t-1} - \omega_0) + \varepsilon_t \quad (3.14)$$

for $t \in \mathbb{Z}$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ is a parameter vector in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. We denote by $\theta = (\omega, \alpha, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 3.3. (*ARMA(1,1)-model*)

- (i) (*Compactness*) Θ is compact;
- (ii) (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;
- (iii) (*Invertibility*) $|\alpha| < 1$ for all $\theta \in \Theta$;
- (iv) (*Causality*) $|\beta| < 1$ for all $\theta \in \Theta$;
- (v) (*Roots*) $1 - \beta z$ and $1 + \alpha z$ have no common root, and $\alpha, \beta \neq 0$ for all $\theta \in \Theta$;
- (vi) (*Innovations*) ε_t are i.i.d. from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = \sigma_\varepsilon^2 < \infty$ and having a Lebesgue density strictly positive on \mathbb{R} ;

Θ is assumed to be compact in Assumption 3.3.(i), which holds true, for instance, if it is of the form $\Theta = \{(\omega, \alpha, \beta)' \in \mathbb{R}^3 : |\omega| \leq \delta^{-1}, \delta \leq |\alpha| \leq 1 - \delta \text{ and } \delta \leq |\beta| \leq 1 - \delta\}$, where $\delta > 0$ is a sufficiently small constant. Assumption 3.3.(ii) states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The invertibility and causality conditions are stated in 3.3.(iii) and 3.3.(iv). Assumption 3.3.(v) ensures that the ARMA model is irreducible. Assumption 3.3.(vi) imposes further

restrictions on the distribution of the innovation process. Next, we turn to the estimation of the model.

3.5.2 Estimation

To estimate the model in equation (3.14), we consider a least squares estimator in the spirit of Brockwell and Davis (1991).³ Other estimators such as the QML estimator based on the Gaussian likelihood can alternatively be considered. Let $G_n(\alpha_0, \beta_0)$ be the correlation matrix of $(X_1, \dots, X_n)'$ with elements given by

$$\text{Cor}(X_t, X_{t-k}) = \frac{(\alpha_0 + \beta_0)(1 + \alpha_0\beta_0)}{1 + 2\alpha_0\beta_0 + \alpha_0^2} \beta_0^{k-1}$$

for $k \geq 1$. The (weighted) least squares estimator of θ_0 is given by

$$\hat{\theta}(\mathbf{X}_{1:n}) = \arg \min_{\theta \in \Theta} (\mathbf{X}_{1:n} - \omega \iota_n)' G_n^{-1}(\alpha, \beta) (\mathbf{X}_{1:n} - \omega \iota_n). \quad (3.15)$$

with $\iota_n = (1, \dots, 1)' \in \mathbb{R}^n$. As the sample size grows large, the estimator approaches a normal distribution under regularity conditions.

Theorem 3.3. (*Brockwell and Davis, 1991; Bao, 2018*) *Under Assumption 3.3*

$$\sqrt{n}(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0) \quad (3.16)$$

with

$$\Upsilon_0 = \begin{pmatrix} \frac{\sigma_\varepsilon^2(1-\alpha_0)^2}{(1-\beta_0)^2} & 0 & 0 \\ 0 & \frac{(1-\alpha_0\beta_0)^2(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} & \frac{(1-\alpha_0^2)(1-\alpha_0\beta_0)(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} \\ 0 & \frac{(1-\alpha_0^2)(1-\alpha_0\beta_0)(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} & \frac{(1-\alpha_0\beta_0)^2(1-\alpha_0^2)}{(\alpha_0-\beta_0)^2} \end{pmatrix}. \quad (3.17)$$

It is worth highlighting that Υ_0 does not only depend on $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$, but also on the nuisance parameter σ_ε^2 .

³Brockwell and Davis (1991) consider $\omega = 0$ for simplicity. The extension to $\omega \neq 0$ is straight-forward.

3.5.3 Mapping

Having described the model and its estimation, we write the model in terms of the general framework. The conditional mean of X_{n+1} is equal to

$$\begin{aligned}\psi_{n+1} &= \psi(X_n, X_{n-1}, \dots; \theta_0) \\ &= \omega_0 + \sum_{k=0}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) (X_{n-k} - \omega_0).\end{aligned}\tag{3.18}$$

The first and second derivatives of $\psi(X_n, X_{n-1}, \dots; \theta)$ w.r.t. θ are need to verify Assumption 2.1. The first order derivatives are

$$\begin{aligned}\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega} &= \frac{1 - \beta}{1 + \alpha}, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=0}^{\infty} (-\alpha)^k (X_{n-k} - \omega), \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha} &= \sum_{k=0}^{\infty} (k+1) (-\alpha)^k \left((X_{n-k} - \omega) - \beta (X_{n-k-1} - \omega) \right),\end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned}\frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} &= 0, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} &= 0, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} &= -\frac{1}{1 + \alpha}, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha} &= -\frac{1 - \beta}{(1 + \alpha)^2}, \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} &= \sum_{k=0}^{\infty} (k+1) (-\alpha)^k (\omega - X_{n-k-1}), \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} &= -\sum_{k=1}^{\infty} (k+1) k (-\alpha)^{k-1} \left((X_{n-k} - \omega) \right. \\ &\quad \left. - \beta (X_{n-k-1} - \omega) \right).\end{aligned}$$

3.5.4 Verification of Assumptions

Before turning to the verification of the assumptions of Chapter 2, note that $\mathbb{E}|X_t| < \infty$ as the process $\{X_t\}$ is assumed to be causal and $\mathbb{E}|\varepsilon_t| < \infty$.⁴

Assumption 2.1

For Assumption 2.1.(i) to be met, we consider the least squares estimator in equation (3.15), whose asymptotic distribution is specified in Theorem 3.3.

As the function $\psi_{n+1}(\dots; \theta)$ given in (3.18) is continuous on Θ and twice differentiable on $\dot{\Theta}$, Assumption 2.1.(ii) is met.

Regarding Assumption 2.1.(iii) we note that

$$\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} = \frac{1 - \beta_0}{1 + \alpha_0}$$

is trivially $O(1)$. To show $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta}\right| \geq M\right] < \epsilon$ for sufficiently large n . Employing the Markov inequality, we obtain

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta}\right| \geq M\right] &\leq \frac{1}{M} \mathbb{E}\left[\left|\sum_{k=0}^{\infty} (-\alpha_0)^k (X_{n-k} - \omega_0)\right|\right] \\ &\leq \frac{1}{M} \sum_{k=0}^{\infty} |\alpha_0|^k (\mathbb{E}|X_t| + |\omega_0|) = \frac{\mathbb{E}|X_t| + |\omega_0|}{(1 - |\alpha_0|)M} \end{aligned}$$

⁴As $\{X_t\}$ is causal, we can write it in the MA(∞) representation: $X_t - \omega_0 = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} |\vartheta_j| < \infty$ such that $\mathbb{E}|X_t| \leq |\omega_0| + \mathbb{E}|\varepsilon_t| \sum_{j=0}^{\infty} |\vartheta_j| < \infty$.

such that $M > \frac{\mathbb{E}|X_t| + |\omega_0|}{(1 - |\alpha_0|)\epsilon}$ gives the desired result. Similarly, we find

$$\begin{aligned}
 & \mathbb{P} \left[\left| \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha} \right| \geq M \right] \\
 & \leq \frac{1}{M} \mathbb{E} \left[\left| \sum_{k=0}^{\infty} (k+1)(-\alpha_0)^k \left((X_{n-k} - \omega_0) - \beta_0(X_{n-k-1} - \omega_0) \right) \right| \right] \\
 & \leq \frac{1}{M} \mathbb{E} \left[\sum_{k=0}^{\infty} (k+1)|\alpha_0|^k \left(|X_{n-k}| + |\omega_0| + |\beta_0| (|X_{n-k-1}| + |\omega_0|) \right) \right] \\
 & = \frac{1}{M} \sum_{k=0}^{\infty} (k+1)|\alpha_0|^k (\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|) = \frac{(\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|)}{M(1 - |\alpha_0|)^2}
 \end{aligned}$$

such that $M > \frac{(\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|)}{\epsilon(1 - |\alpha_0|)^2}$ establishes $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha} = O_p(1)$, which completes the verification of Assumption 2.1.(iii).

Consider Assumption 2.1.(iv) and note that

$$\begin{aligned}
 & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} \right| = 0, \\
 & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} \right| = 0, \\
 & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} \right| = \frac{1}{1 - \alpha_{\sup}} = O(1)
 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha} \right| \leq \frac{1 + \beta_{\sup}}{1 - \alpha_{\sup}},$$

where $\alpha_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} |\alpha|$ as well as $\beta_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} |\beta|$. To show that the term $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right|$ is $O_p(1)$, we need to find an M for every $\epsilon > 0$ such that $\mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] < \epsilon$ holds for sufficiently large n . We obtain

$$\begin{aligned}
 & \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] \\
 & \leq \frac{1}{M} \mathbb{E} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=0}^{\infty} (k+1)(-\alpha)^k (\omega - X_{n-k-1}) \right| \right]
 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{1}{M} \mathbb{E} \left[\sum_{k=0}^{\infty} (k+1) \alpha_{\sup}^k (|X_{n-k-1}| + \omega_{\sup}) \right] \\
 &\leq \frac{1}{M} \sum_{k=0}^{\infty} (k+1) \alpha_{\sup}^k (\mathbb{E}|X_t| + \omega_{\sup}) = \frac{\mathbb{E}|X_t| + \omega_{\sup}}{M(1 - \alpha_{\sup})^2},
 \end{aligned}$$

where $\omega_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} |\omega|$. Taking $M > \frac{\mathbb{E}|X_t| + \omega_{\sup}}{\epsilon(1 - \alpha_{\sup})^2}$ leads to the desired result. Similarly, we find

$$\begin{aligned}
 &\mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} \right| \geq M \right] \\
 &\leq \frac{1}{M} \mathbb{E} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=1}^{\infty} (k+1) k (-\alpha)^{k-1} \left((X_{n-k} - \omega) - \beta(X_{n-k-1} - \omega) \right) \right| \right] \\
 &\leq \frac{1}{M} \mathbb{E} \left[\sum_{k=1}^{\infty} (k+1) k \alpha_{\sup}^{k-1} \left(|X_{n-k}| + \omega_{\sup} + \beta_{\sup} (|X_{n-k-1}| + \omega_{\sup}) \right) \right] \\
 &\leq \frac{1}{M} \sum_{k=1}^{\infty} (k+1) k \alpha_{\sup}^{k-1} (\mathbb{E}|X_t| + \omega_{\sup}) (1 + \beta_{\sup}) = \frac{2(\mathbb{E}|X_t| + \omega_{\sup})(1 + \beta_{\sup})}{M(1 - \alpha_{\sup})^3}.
 \end{aligned}$$

Taking $M > \frac{2(\mathbb{E}|X_t| + \omega_{\sup})(1 + \beta_{\sup})}{\epsilon(1 - \alpha_{\sup})^3}$ establishes that $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} \right| = O_p(1)$.

Regarding Assumption 2.1.(v) we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e. $c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_n, X_{n-1}, \dots; \theta) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta) = \sum_{k=n-t_1+1}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) X_{n-k}.$$

We have

$$\begin{aligned}
 &m_n \left(\psi(X_n, X_{n-1}, \dots; \theta_0) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) \right) \\
 &= \sqrt{n} (-\alpha_0)^{n-t_1} \sum_{k=1}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) X_{t_1-k}.
 \end{aligned} \tag{3.19}$$

Clearly, the sum is of order $O_p(1)$ as $|\alpha_0| < 1$ and $\{X_t\}$ is strictly stationary. Further, for any $t_1 \geq 1$ such that $(n - t_1)/l_n \rightarrow \infty$ we get $\sqrt{n}(-\alpha_0)^{n-t_1} \rightarrow 0$.

Hence, (3.19) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

and

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} \right| \\ &= \left| \sum_{k=n-t_1+1}^{\infty} (-\alpha_0)^k X_{n-k} \right| \leq |\alpha_0|^{n-t_1} \sum_{k=1}^{\infty} |\alpha_0|^k |X_{t_1-k}| \end{aligned}$$

being $o_p(1)$ since the sum is $O_p(1)$ and $|\alpha_0|^{n-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \alpha} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha} \right| \\ &= \left| \sum_{k=n-t_1+1}^{\infty} ((k+1)\alpha_0 + k\beta_0)(-\alpha_0)^{k-1} X_{n-k} \right| \\ &\leq \sum_{k=n-t_1+1}^{\infty} ((k+1)|\alpha_0| + k|\beta_0|)|\alpha_0|^{k-1} |X_{n-k}| \\ &\leq 2(|\alpha_0| + |\beta_0|) \sum_{k=n-t_1+1}^{\infty} k|\alpha_0|^{k-1} |X_{n-k}| \\ &= (n-t_1)|\alpha_0|^{n-t_1} 2(|\alpha_0| + |\beta_0|) \sum_{k=1}^{\infty} |\alpha_0|^{k-1} |X_{t_1-k}| \\ &\quad + |\alpha_0|^{n-t_1} 2(|\alpha_0| + |\beta_0|) \sum_{k=1}^{\infty} k|\alpha_0|^{k-1} |X_{t_1-k}| \end{aligned}$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| = (0, 0, 0), \\ & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \beta^2} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} \right| = 0 \end{aligned}$$

and

$$\begin{aligned}
 & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha \partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \\
 &= \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=n-t_1}^{\infty} (k+1)(-\alpha)^k X_{n-k-1} \right| \leq \sum_{k=n-t_1}^{\infty} (k+1) \alpha_{\sup}^k |X_{n-k-1}| \\
 &= (n-t_1) \alpha_{\sup}^{n-t_1} \sum_{k=0}^{\infty} \alpha_{\sup}^k |X_{t_1-k-1}| + \alpha_{\sup}^{n-t_1} \sum_{k=0}^{\infty} (k+1) \alpha_{\sup}^k |X_{t_1-k-1}|
 \end{aligned}$$

is $o_p(1)$ by previous arguments as $\alpha_{\sup} \in (0, 1)$. Similarly, it can be shown that

$$\begin{aligned}
 & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^2} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^2} \right| \\
 &= \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=n-t_1+1}^{\infty} (k+1)k(-\alpha)^{k-1} X_{n-k} \right. \\
 & \quad \left. - \beta \sum_{k=n-t_1}^{\infty} (k+1)k(-\alpha)^{k-1} X_{n-k-1} \right| \\
 &\leq \sum_{k=n-t_1+1}^{\infty} (k+1)k \alpha_{\sup}^{k-1} |X_{n-k}| + \beta_{\sup} \sum_{k=n-t_1}^{\infty} (k+1)k \alpha_{\sup}^{k-1} |X_{n-k-1}|
 \end{aligned}$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 2.3

The condition in Assumption 2.3.(i) is satisfied for instance by $n_E(n) \sim n - \lfloor n^b \rfloor$ and $n_P(n) \sim n - \lfloor n^a \rfloor$ with $0 < a < b < 1$.

The process $\{X_t\}$ is strictly stationary since $|\beta_0| < 1$ and $\mathbb{E} \log^+ |\varepsilon_t| \leq \mathbb{E} |\varepsilon_t| < \infty$ (Bougerol and Picard, 1992, Theorem 4.1).

The process $\{(\varepsilon_t, X_t)\}$ is β -mixing with exponential decay (Mokkadem, 1988, Theorem 1'). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 2.3.(iii) is met with regard to Remark 2.3 noting that $n_P(n) - n_E(n) \rightarrow \infty$. For an alternative mixing result we refer to Davidson (1994, Theorem 14.9).

Assumptions 2.4 and 2.5

Assumption 2.4 is implied by Assumption 2.5, which, in turn, is verified by Theorem 3.3 and the consistent estimator

$$\hat{\Upsilon}(\mathbf{X}_{1:n}) = \begin{pmatrix} \hat{\Upsilon}_{11}(\mathbf{X}_{1:n}) & 0 & 0 \\ 0 & \hat{\Upsilon}_{22}(\mathbf{X}_{1:n}) & \hat{\Upsilon}_{23}(\mathbf{X}_{1:n}) \\ 0 & \hat{\Upsilon}_{23}(\mathbf{X}_{1:n}) & \hat{\Upsilon}_{33}(\mathbf{X}_{1:n}) \end{pmatrix},$$

where

$$\begin{aligned} \hat{\Upsilon}_{11}(\mathbf{X}_{1:n}) &= \frac{\hat{\sigma}_\varepsilon^2(\mathbf{X}_{1:n})(1 - \hat{\alpha}(\mathbf{X}_{1:n}))^2}{(1 - \hat{\beta}(\mathbf{X}_{1:n}))^2} \\ \hat{\Upsilon}_{22}(\mathbf{X}_{1:n}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:n})\hat{\beta}(\mathbf{X}_{1:n}))^2(1 - \hat{\beta}(\mathbf{X}_{1:n})^2)}{(\hat{\alpha}(\mathbf{X}_{1:n}) - \hat{\beta}(\mathbf{X}_{1:n}))^2} \\ \hat{\Upsilon}_{23}(\mathbf{X}_{1:n}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:n})^2)(1 - \hat{\alpha}(\mathbf{X}_{1:n})\hat{\beta}(\mathbf{X}_{1:n}))(1 - \hat{\beta}(\mathbf{X}_{1:n})^2)}{(\hat{\alpha}(\mathbf{X}_{1:n}) - \hat{\beta}(\mathbf{X}_{1:n}))^2} \\ \hat{\Upsilon}_{33}(\mathbf{X}_{1:n}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:n})\hat{\beta}(\mathbf{X}_{1:n}))^2(1 - \hat{\alpha}(\mathbf{X}_{1:n})^2)}{(\hat{\alpha}(\mathbf{X}_{1:n}) - \hat{\beta}(\mathbf{X}_{1:n}))^2} \end{aligned}$$

and $\hat{\sigma}_\varepsilon^2(\mathbf{X}_{1:n})$ equals to

$$\frac{1}{n-3}(\mathbf{X}_{1:n} - \hat{\omega}(\mathbf{X}_{1:n})\boldsymbol{\iota}_n)' G_n^{-1}(\hat{\alpha}(\mathbf{X}_{1:n}), \hat{\beta}(\mathbf{X}_{1:n}))(\mathbf{X}_{1:n} - \hat{\omega}(\mathbf{X}_{1:n})\boldsymbol{\iota}_n).$$

Assumptions within Corollary 2.1

To show $1/\hat{v}_n^{2IP} = O_p(1)$, recall that $\hat{v}_n^{2IP} = v_n^{2IP} + o_p(1)$ (see proof of Corollary 2.2) and define $\kappa = \text{eig}_{\min} \Upsilon_0$, the minimum eigenvalue of Υ_0 . Since Υ_0 is positive definite, we have $\kappa > 0$. Together with $|\beta_0| < 1$ and $|\alpha_0| < 1$

$$\begin{aligned} \hat{v}_n^{2IP} + o_p(1) &= v_n^{2IP} \geq \kappa \left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta_0)}{\partial \theta} \right\|^2 \\ &\geq \kappa \left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{1:n}; \theta_0)}{\partial \omega} \right\|^2 = \kappa \frac{(1 - \beta_0)^2}{(1 + \alpha_0)^2} > 0 \end{aligned}$$

implies $1/\hat{v}_n^{2IP} = O_p(1)$. Analogously, we have $1/\hat{v}_n^{SPL} = O_p(1)$.

3.6 Conditional Volatility in a T-GARCH(1,1)

3.6.1 Model Description

The threshold GARCH (T-GARCH) model was first introduced by Zakoïan (1994). It accounts for the stylized fact that past positive and negative innovations appear not to have the same impact on current volatility, which is also known as leverage effect. The T-GARCH(1, 1) process $\{X_t\}$ is defined by

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t &= \omega_0 + \alpha_0^+ X_{t-1}^+ + \alpha_0^- X_{t-1}^- + \beta_0 \sigma_{t-1} \end{aligned} \tag{3.20}$$

for all $t \in \mathbb{Z}$ using the notation $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)'$ are non-negative parameters in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. We denote by $\theta = (\omega, \alpha^+, \alpha^-, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 3.4. (*T-GARCH(1,1)-model*)

- (i) (*Compactness*) Θ is compact;
- (ii) (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;
- (iii) (*Non-negativity*) $\omega > 0$, $\alpha^+ \geq 0$, $\alpha^- \geq 0$ and $\beta \geq 0$ for all $\theta \in \Theta$;
- (iv) (*Strict stationarity*) $\mathbb{E}[\ln(\alpha_0^+ \varepsilon_t^+ + \alpha_0^- \varepsilon_t^- + \beta_0)] < 1$ and $\beta < 1$ for all $\theta \in \Theta$;
- (v) (*Roots*) $1 - \beta_0 z > 0$ has no common root with $\alpha_0^+ z$ and $\alpha_0^- z$, and $\alpha_0^+ + \alpha_0^- \neq 0$;
- (vi) (*Innovations*) ε_t are i.i.d. from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = 1$ and $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive in a neighborhood of zero.

Θ is assumed to be compact in Assumption 3.4.(i), which holds true, for instance, if it is of the form $\Theta = [\delta, 1/\delta] \times [0, 1/\delta]^2 \times [0, 1 - \delta]$, where $\delta \in (0, 1)$ is a sufficiently small constant. Assumption 3.4.(ii) states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The non-negativity constraints in 3.4.(iii) are standard ensuring the conditional standard deviation to be strictly positive.

Assumption 3.4.(iv) is necessary and sufficient for $\{X_t\}$ being strictly stationary (cf. Hamadeh and Zakoïan, 2011). The root condition in 3.4.(v) guarantees that the T-GARCH model is irreducible. Assumption 3.4.(vi) imposes further restrictions on the moments and density of the innovation process. Next, we turn to the estimation of the model in (3.20).

3.6.2 Estimation

We consider the Gaussian QML estimator proposed by Hamadeh and Zakoïan (2011). For a generic $\theta \in \Theta$ we set

$$\sigma_{t+1}(\theta) = \sum_{k=0}^{\infty} \beta^k (\omega + \alpha^+ X_{t-k}^+ + \alpha^- X_{t-k}^-) \quad (3.21)$$

and note that $\sigma_{t+1} = \sigma_{t+1}(\theta_0)$. Replacing the unknown presample observations by arbitrary values, say s_t , $t \leq 0$, we denote the modified version of (3.21) by $\tilde{\sigma}_{t+1}^2(\theta)$. Then the QML estimator of θ_0 is defined as any measurable solution $\hat{\theta}(\mathbf{X}_{1:n})$ of

$$\hat{\theta}(\mathbf{X}_{1:n}) = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta; \mathbf{X}_{1:n}) \quad (3.22)$$

with

$$\tilde{L}_n(\theta; \mathbf{X}_{1:n}) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\theta)}} \exp\left(-\frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)}\right).$$

Assumption 3.4 implies that the estimator follows asymptotically a normal distribution.

Theorem 3.4. (*Hamadeh and Zakoïan, 2011*) *Under Assumption 3.4*

$$\sqrt{n}(\hat{\theta}(\mathbf{X}_{1:n}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0), \quad (3.23)$$

where $\Upsilon_0 = \frac{1}{4}(\mathbb{E}[\varepsilon_t^4] - 1) \mathbb{E}\left[\frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'}\right]^{-1}$ and $\sigma_t(\theta)$ is given in (3.21).

3.6.3 Mapping

Having described the model and its estimation, we map the model into the general framework. The conditional volatility σ_{n+1} is equal to

$$\psi_{n+1} = \psi(X_n, X_{n-1}, \dots; \theta_0) = \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{n-k}^+ + \alpha_0^- X_{n-k}^-). \quad (3.24)$$

To verify Assumption 2.1 the first and second derivatives of $\psi(X_n, X_{n-1}, \dots; \theta)$ w.r.t. θ are needed. The first order derivatives are

$$\begin{aligned} \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega} &= \frac{1}{1 - \beta}, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+} &= \sum_{k=0}^{\infty} \beta^k X_{n-k}^+, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^-} &= \sum_{k=0}^{\infty} \beta^k X_{n-k}^-, \\ \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} (\omega + \alpha^+ X_{n-k}^+ + \alpha^- X_{n-k}^-), \end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned} \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha^+} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha^-} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{+2}} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \alpha^-} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{-2}} &= 0 \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} &= \frac{1}{(1 - \beta)^2} \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{n-k}^+ \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{n-k}^- \\ \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} &= \sum_{k=2}^{\infty} k(k-1) \beta^{k-2} (\omega + \alpha^+ X_{n-k}^+ + \alpha^- X_{n-k}^-)\end{aligned}$$

3.6.4 Verification of Assumptions

Before turning to the verification of the assumptions of Chapter 2, note that the strict stationarity condition implies the existence of fractional moments: there exists an $s \in (0, 1)$ such that $E|X_t|^s < \infty$ (Hamadeh and Zakoïan, 2011, Proposition A.1).

Assumption 2.1

For Assumption 2.1.(i) to be met, we consider the quasi-maximum likelihood estimator by Hamadeh and Zakoïan (2011), whose asymptotic distribution is specified in Theorem 3.4.

As the function $\psi(\dots; \theta)$, given in (3.24), is continuous on Θ and twice differentiable on $\mathring{\Theta}$, Assumption 2.1.(ii) is satisfied.

Consider Assumption 2.1.(iii) and note that

$$\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} = \frac{1}{1 - \beta_0}$$

is trivially $O(1)$. For showing $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^+} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^+}\right| \geq M\right] < \epsilon$ for n sufficiently large. Markov's inequality implies

$$\begin{aligned}\mathbb{P}\left[\left|\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^+}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{n-k}^+\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k |X_{n-k}| \right)^s\right] \leq \frac{1}{M^s} \sum_{k=0}^{\infty} \beta_0^{sk} \mathbb{E}|X_t|^s = \frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s) M^s}\end{aligned}$$

such that $M > \left(\frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s)\epsilon}\right)^{1/s}$ gives the desired result. The same M serves to show

$$\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^-} = O_p(1):$$

$$\begin{aligned} \mathbb{P} \left[\left| \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^-} \right| \geq M \right] &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{n-k}^- \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=0}^{\infty} \beta_0^k |X_{n-k}| \right)^s \right] \leq \frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s) M^s} < \epsilon. \end{aligned}$$

Similarly, we get

$$\begin{aligned} &\mathbb{P} \left[\left| \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} \right| \geq M \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_0^{k-1} (\omega_0 + \alpha_0^+ X_{n-k}^+ + \alpha_0^- X_{n-k}^-) \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k \beta_0^{s(k-1)} (\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s |X_{n-k}|^s) \right] = \frac{\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s \mathbb{E}|X_t|^s}{M^s (1 - \beta_0^s)^2} \end{aligned}$$

such that $M > \left(\frac{\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s \mathbb{E}|X_t|^s}{\epsilon (1 - \beta_0^s)^2} \right)^{1/s}$ establishes $\frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} = O_p(1)$.

Concerning Assumption 2.1.(iv) we notice that

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega^2} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha^+} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \alpha^-} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{+2}} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \alpha^-} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{-2}} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \beta} \right| = \frac{1}{(1 - \beta_{\sup})^2} = O(1),$$

where $\beta_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \beta$. To show $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| = O_p(1)$, we need to find an M for every $\epsilon > 0$ such that $\mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| \geq M \right] < \epsilon$ holds. We obtain

$$\begin{aligned} \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| \geq M \right] &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^+ \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\sup}^{s(k-1)} |X_{n-k}|^s \right] \leq \frac{1}{M^s} \sum_{k=1}^{\infty} k \beta_{\sup}^{s(k-1)} \mathbb{E} |X_t|^s = \frac{\mathbb{E} |X_t|^s}{M^s (1 - \beta_{\sup}^s)^2}. \end{aligned}$$

Taking $M > \left(\frac{\mathbb{E} |X_t|^s}{\epsilon (1 - \beta_{\sup}^s)^2} \right)^{1/s}$ leads to the desired result. The same M serves to prove that $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| = O_p(1)$ since

$$\begin{aligned} \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| \geq M \right] &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^- \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\sup}^{s(k-1)} |X_{n-k}|^s \right] \leq \frac{\mathbb{E} |X_t|^s}{M^s (1 - \beta_{\sup}^s)^2} < \epsilon. \end{aligned}$$

Similarly, we have

$$\begin{aligned} &\mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \beta^2} \right| \geq M \right] \\ &= \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=2}^{\infty} k(k-1) \beta_{\sup}^{k-2} (\omega_{\sup} + \alpha_{\sup}^+ X_{n-k}^+ + \alpha_{\sup}^- X_{n-k}^-) \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=2}^{\infty} k^s (k-1)^s \beta_{\sup}^{s(k-2)} (\omega_{\sup}^s + (\alpha_{\sup}^+ + \alpha_{\sup}^-)^s |X_{n-k}|^s) \right] \\ &\leq \frac{1}{M^s} \sum_{k=2}^{\infty} k(k-1) \beta_{\sup}^{s(k-2)} (\omega_{\sup}^s + (\alpha_{\sup}^+ + \alpha_{\sup}^-)^s \mathbb{E} |X_t|^s) \\ &= \frac{2(\omega_{\sup}^s + (\alpha_{\sup}^+ + \alpha_{\sup}^-)^s \mathbb{E} |X_t|^s)}{M^s (1 - \beta_{\sup}^s)^3}, \end{aligned}$$

where $\omega_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \omega$, $\alpha_{\sup}^+ = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha^+$ and $\alpha_{\sup}^- = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha^-$. Taking $M > \left(\frac{2(\omega_{\sup}^s + (\alpha_{\sup}^+ + \alpha_{\sup}^-)^s \mathbb{E} |X_t|^s)}{\epsilon (1 - \beta_{\sup}^s)^3} \right)^{1/s}$ completes the verification of Assumption 2.1.(iv).

Regarding Assumption 2.1.(v) we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e. $c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_n, X_{n-1}, \dots; \theta) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta) = \sum_{k=n-t_1+1}^{\infty} \beta^k (\alpha^+ X_{n-k}^+ + \alpha^- X_{n-k}^-).$$

We have

$$\begin{aligned} m_n \left(\psi(X_n, X_{n-1}, \dots; \theta_0) - \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0) \right) \\ = \sqrt{n} \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^k (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-). \end{aligned} \quad (3.25)$$

Clearly, the sum is of order $O_p(1)$. Further, for any $t_1 \geq 1$ such that $(n - t_1)/l_n \rightarrow \infty$ we get $\sqrt{n} \beta_0^{n-t_1} \rightarrow 0$. Hence, (3.25) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

as well as

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \alpha^+} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^+} \right| \\ &= \sum_{k=n-t_1+1}^{\infty} \beta_0^k X_{n-k}^+ \leq \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^+ \end{aligned}$$

and

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \alpha^-} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \alpha^-} \right| \\ &= \sum_{k=n-t_1+1}^{\infty} \beta_0^k X_{n-k}^- \leq \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^- \end{aligned}$$

being $o_p(1)$ since the sums are $O_p(1)$ and $\beta_0^{n-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned} & \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta} \right| \\ &= \sum_{k=n-t_1+1}^{\infty} k \beta_0^{k-1} (\alpha_0^+ X_{n-k}^+ + \alpha_0^- X_{n-k}^-) \\ &= (n - t_1) \beta_0^{n-t_1} \sum_{k=1}^{\infty} \beta_0^{k-1} (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-) \end{aligned}$$

$$+ \beta_0^{n-t_1} \sum_{k=1}^{\infty} k \beta_0^{k-1} (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-)$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| &= (0, 0, 0), \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^{+2}} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{+2}} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^{-2}} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^{-2}} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^+ \partial \alpha^-} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \alpha^-} \right| = 0.$$

In addition, we find

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^+ \partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| &= \sum_{k=n-t_1+1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^+ \\ &= (n-t_1) \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} \beta_{\sup}^{k-1} X_{t_1-k}^+ + \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{t_1-k}^+ \end{aligned}$$

and

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \alpha^- \partial \beta} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| &= \sum_{k=n-t_1+1}^{\infty} k \beta_{\sup}^{k-1} X_{n-k}^- \\ &= (n-t_1) \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} \beta_{\sup}^{k-1} X_{t_1-k}^- + \beta_{\sup}^{n-t_1} \sum_{k=1}^{\infty} k \beta_{\sup}^{k-1} X_{t_1-k}^- \end{aligned}$$

being $o_p(1)$ by previous arguments noting that $\beta_{\sup} \in (0, 1)$. Similarly, it can be shown that

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta_0)}{\partial \beta \partial \beta'} - \frac{\partial \psi(X_n, X_{n-1}, \dots; \theta_0)}{\partial \beta \partial \beta'} \right|$$

$$\leq \sum_{k=n-t_1+1}^{\infty} k(k-1)\beta_{\sup}^{k-2}(\alpha_{\sup}^+ X_{n-k}^+ + \alpha_{\sup}^- X_{n-k}^-)$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{n+1}^s(\mathbf{X}_{t_1:n}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_n, X_{n-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 2.3

The condition in Assumption 2.3.(i) is satisfied for instance by $n_E(n) \sim n - \lfloor n^b \rfloor$ and $n_P(n) \sim n - \lfloor n^a \rfloor$ with $0 < a < b < 1$.

With regard to Assumption 3.4.(iv), $\{X_t\}$ is a strictly stationary process such that Assumption 2.3.(ii) is satisfied.

The process $\{X_t\}$ is β -mixing with exponential decay (Francq and Zakoian, 2006, Theorem 3). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 2.3.(iii) is met with regard to Remark 2.3 noting that $n_P(n) - n_E(n) \rightarrow \infty$. For an alternative mixing result we refer to Carrasco and Chen (2002).

Assumptions 2.4 and 2.5

Assumption 2.4 is implied by Assumption 2.5, which, in turn, is verified by Theorem 3.4 and the consistent estimator

$$\begin{aligned} \hat{\Upsilon}(\mathbf{X}_{1:n}) &= \left(\frac{1}{n} \sum_{t=1}^n \frac{X_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:n}))} - 1 \right) \\ &\quad \times \left(\frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:n}))} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:n}))}{\partial \theta'} \right)^{-1}. \end{aligned}$$

Assumptions within Corollary 2.1

The verification of $1/\hat{v}_n^{2IP} = O_p(1)$ and $1/\hat{v}_n^{SPL} = O_p(1)$ is analogous to the GARCH(1,1) case and hence omitted.

Model	Conditional variance σ_t^2 specification
E-GARCH(1, 1)	$\ln \sigma_t^2 = \omega + \alpha \frac{X_{t-1}}{\sigma_{t-1}} + \phi \left(\left \frac{X_{t-1}}{\sigma_{t-1}} \right - \mathbb{E} \varepsilon_t \right) X_{t-1}^2 + \beta \ln \sigma_{t-1}^2$
N-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha (X_{t-1} - \phi \sigma_{t-1})^2 + \beta \sigma_{t-1}^2$
GJR-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha^+ X_{t-1}^2 I_{\{X_{t-1} \geq 0\}} + \alpha^- X_{t-1}^2 I_{\{X_{t-1} < 0\}} + \beta \sigma_{t-1}^2$
Q-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 + \phi X_{t-1}$

Table 3.1: GARCH extensions. The respective process $\{X_t\}$ is generated by $X_t = \sigma_t \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of innovations and σ_t^2 is the conditional variance at time t .

3.7 Concluding Remarks

In this chapter we establish the mapping of the conditional mean in an AR(1) and ARMA(1, 1) model into the general framework. Further, the conditional variance and the conditional volatility in a GARCH(1, 1) and T-GARCH(1, 1) model, respectively, are shown to be encompassed in this setup. Further, the theoretical results of Chapter 2 are validated by verifying the corresponding assumptions for each model. Clearly, the list of nested models is non-exhaustive and can be extended. For instance one could study higher order models such as the ARMA(p, q) or the GARCH(p, q) model with $p, q \in \mathbb{N}$, which come at the cost of a more evolved analysis. Table 3.1 enlists four other GARCH-type extensions that are frequently encountered in the literature. The family of quadratic GARCH (Q-GARCH) models has been proposed by Sentana (1995). Its Q-GARCH(1, 1) member is very similar to the GARCH(1, 1) model and can be verified in a similar fashion replacing αX_{t-1}^2 by $\alpha X_{t-1}^2 + \phi X_t$. The GJR-GARCH(1, 1) model named after Glosten, Jagannathan and Runkle (1993) is a variant of the T-GARCH(1, 1), which corresponds to squaring the variables involved. It can be easily verified along the lines of Section 3.6. The exponential GARCH (E-GARCH) model suggested by Nelson (1991) and the non-linear GARCH (N-GARCH) introduced by Engle and Ng (1993) can also be embedded into the general framework. For example, the conditional variance in an N-GARCH(1, 1) given by $\sigma_{n+1}^2 = \omega_0 + \alpha_0 (X_n - \phi_0 \sigma_n)^2 + \beta_0 \sigma_n^2$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0, \phi_0)'$ denotes the parameter vector. However, obtaining an

explicit expression for the conditional variance in terms of θ_0 and $\{X_t\}_{t \leq n}$ is complicated due to non-linearities in the recursive formula: e.g. σ_{n+1}^2 depends on σ_n^2 and σ_n in the N-GARCH(1,1).

There are few GARCH extensions such as the fractionally integrated (FI-GARCH) of Baillie, Bollerslev, and Mikkelsen (1996) or the fractionally integrated EGARCH (FIE-GARCH) of Bollerslev and Mikkelsen (1996) that cannot be encompassed in the framework at hand. The corresponding processes typically exhibit intermediate or long memory such that standard mixing results do not apply. Establishing the merging results on the basis of verifying Assumption 2.3.(iii) directly, instead via some mixing result, is an interesting question, which demands further investigation.

Finally, we would like to emphasize that conditional risk measures such as conditional Value-at-Risk (VaR) can be mapped into the general framework. For instance in the T-GARCH(1,1) model of Section 3.6, the conditional VaR of X_{n+1} given $\{X_t\}_{t \leq n}$ at level $a \in (0, 1)$ reduces to

$$\begin{aligned} & VaR_a(X_{n+1}|X_n, X_{n-1}, \dots) \\ &= -\xi_a \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{n-k}^+ + \alpha_0^- X_{n-k}^-) \end{aligned} \quad (3.26)$$

with $\xi_a = \inf \{ \tau \in \mathbb{R} : \mathbb{P}[\varepsilon_t \leq \tau] \geq a \}$; see Francq and Zakoïan (2015) for details. Fixing a and treating ξ_a as additional parameter, (3.26) is a function of $\{X_t\}_{t \leq n}$ and $\vartheta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \xi_a)'$ and hence is nested in the setup. Similarly, the conditional Expected Shortfall (ES) of X_{n+1} given $\{X_t\}_{t \leq n}$ at level $a \in (0, 1)$

$$\begin{aligned} & ES_a(X_{n+1}|X_n, X_{n-1}, \dots) \\ &= -\mu_a \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{n-k}^+ + \alpha_0^- X_{n-k}^-) \end{aligned} \quad (3.27)$$

with $\mu_a = -\mathbb{E}[\varepsilon_t | \varepsilon_t < \xi_a]$ can also be mapped into the general framework.

Chapter 4

A Residual Bootstrap for Conditional Value-at-Risk

This chapter proposes a fixed-design residual bootstrap method for the two-step estimator of Francq and Zakoïan (2015) associated with the conditional Value-at-Risk. The bootstrap's consistency is proven under mild assumptions for a general class of volatility models and intervals are constructed for the conditional Value-at-Risk. A simulation study reveals that the equal-tailed percentile bootstrap interval tends to fall short of its nominal value. In contrast, the reversed-tails bootstrap interval yields accurate coverage. We also compare the theoretically analyzed fixed-design bootstrap with the recursive-design bootstrap. It turns out that the fixed-design bootstrap performs equally well in terms of average coverage, yet leads on average to shorter intervals in smaller samples. An empirical application illustrates the interval estimation.¹

¹This chapter is based on the paper Beutner, Heinemann, and Smeekes (2018).

4.1 Introduction

Risk management has tremendously developed in past decades becoming an increasing practice. With minimum capital requirements being enforced by current legislation (Basel III and Solvency II), financial institutions and insurance companies monitor risk by using conventional measures such as Value-at-Risk (VaR). Typically, the volatility dynamics are specified by a (semi-)parametric model leading to conditional risk measure versions. For GARCH-type models the conditional VaR reduces to the conditional volatility scaled by a quantile of the innovations' distribution. The latter is conventionally treated as additional parameter and forms together with the others the *risk parameter* (Francq and Zakoïan, 2015). The true parameters are generally unknown and need to be estimated to obtain an estimate for the conditional VaR. Clearly, this VaR evaluation is subject to estimation risk that needs to be quantified for appropriate risk management.

Whereas an estimator based on a single step is available after reparameterization (Francq and Zakoïan, 2015), a widely used approach is the following two-step estimation procedure. First, the parameters of the conditional volatility model are estimated. Arguably the most popular estimation method in a GARCH-type setting is the Gaussian quasi-maximum-likelihood (QML) method. Based on the model's residuals the quantile is estimated by its empirical counterpart in a second step. For realistic sample sizes (e.g. 500 or 1,000 daily observations) the estimators are subject to considerable estimation risk. In particular, the estimation uncertainty associated with the quantile estimator is substantial for extreme quantiles (e.g. 5% or smaller).

To quantify the uncertainty around the point estimators, one traditionally relies on asymptotic theory while replacing the unknown quantities in the limiting distribution by consistent estimates. An alternative approach – frequently employed in practice – is based on a bootstrap approximation. Regarding the estimators of the GARCH parameters, various bootstrap methods have been studied to approximate the estimators' finite sample distribution including the subsample bootstrap (Hall and Yao, 2003), the block bootstrap (Corradi and Iglesias, 2008), the wild bootstrap (Shimizu, 2010) and the residual bootstrap. The residual bootstrap method is particularly popular and can be further divided into recursive (Pascual et al., 2006; Hidalgo and Zaffaroni, 2007; Jeong, 2017) and fixed (Shimizu, 2010; Cavaliere, Pedersen, and Rahbek, 2018) design. Whereas in the former the bootstrap observations are generated recursively using the estimated volatility dynamics, the

latter design keeps the dynamics of the bootstrap samples fixed at the value of the original series.

The estimation of the quantile and the conditional VaR have received only selected attention in the bootstrap literature and proposed bootstrap methods have been, to the best of our knowledge, exclusively investigated by means of simulation. Christoffersen and Gonçalves (2005) examine various quantile estimators and construct intervals for the conditional VaR using a recursive-design residual bootstrap method. In addition, Hartz, Mittnik, and Paoletta (2006) presume the innovation distribution to be standard normal such that the quantile parameter is known; they propose a resampling method based on a residual bootstrap and a bias-correction step to account for deviations from the normality assumption. In contrast, Spierdijk (2016) develops an m -out-of- n without-replacement bootstrap to construct confidence intervals for ARMA-GARCH VaR.

We propose a fixed-design residual bootstrap method to mimic the finite sample distribution of the two-step estimator and provides an algorithm for the construction of bootstrap intervals for the conditional VaR. The proposed bootstrap method is proven to be consistent for a general class of volatility models. In particular, our framework does not only encompass GARCH but also several GARCH extensions such as the threshold GARCH (T-GARCH) of Zakoïan (1994) and the GJR-GARCH named after Glosten, Jagannathan and Runkle (1993). The bootstrap consistency is established under a set of mild assumptions, which relaxes moment conditions on the innovations imposed in the GARCH bootstrap literature. To the best of our knowledge we are the first to theoretically validate the residual bootstrap for the quantile and the conditional VaR.

The remainder of the chapter is organized as follows. Section 4.2 specifies the model and the conditional VaR is derived. The two-step estimation procedure is described in Section 4.3 and asymptotic theory is provided under mild assumptions. In Section 4.4, a fixed-design residual bootstrap method is proposed and proven to be consistent. Further, bootstrap intervals are constructed for the conditional VaR. A simulation study is conducted in Section 4.5 and an empirical application illustrates the interval estimation based on the fixed-design residual bootstrap. Section 4.6 concludes and auxiliary results are gathered in the Appendix. Appendix 4.A contains lemmas and their proofs while Appendix 4.B is devoted to the related recursive-design residual bootstrap.

4.2 Model

We consider a conditional volatility model of the form

$$\epsilon_t = \sigma_t \eta_t \quad (4.1)$$

with $t \in \mathbb{Z}$, where $\{\epsilon_t\}$ denotes the sequence of log-returns, $\{\sigma_t\}$ is a volatility process and $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) variables. The specification in (4.1) of the observable process $\{\epsilon_t\}$ is particularly suitable for daily financial data. The volatility is presumed to be a measurable function of past observations

$$\sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \quad (4.2)$$

with $\sigma : \mathbb{R}^\infty \times \Theta \rightarrow (0, \infty)$ and θ_0 denotes the true parameter vector belonging to the parameter space $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}$. Subsequently, we consider two examples for the functional form of (4.2): the well-known GARCH model (Engle, 1982; Bollerslev, 1986) and the T-GARCH model of Zakoïan (1994). Whereas the first is frequently applied in practice, the second is motivated by our empirical application (see Section 4.5).

Example 4.1. Suppose $\{\epsilon_t\}$ follows a GARCH(1, 1) process given by (4.1) and

$$\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2,$$

where $\theta_0 = (\omega_0, \alpha_0, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, 1)$. The recursive structure implies

$$\sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) = \sqrt{\sum_{k=1}^{\infty} \beta_0^{k-1} (\omega_0 + \alpha_0 \epsilon_{t-k}^2)}.$$

Example 4.2. Suppose $\{\epsilon_t\}$ follows a T-GARCH(1, 1) process given by (4.1) and

$$\sigma_t = \omega_0 + \alpha_0^+ \epsilon_{t-1}^+ + \alpha_0^- \epsilon_{t-1}^- + \beta_0 \sigma_{t-1}$$

with parameters $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, \infty) \times [0, 1)$ and $\epsilon_t^+ = \max\{\epsilon_t, 0\}$ and $\epsilon_t^- = \max\{-\epsilon_t, 0\}$. The model's recursive structure yields

$$\sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) = \sum_{k=1}^{\infty} \beta_0^{k-1} (\omega_0 + \alpha_0^+ \epsilon_{t-k}^+ + \alpha_0^- \epsilon_{t-k}^-).$$

Throughout the chapter, for any cumulative distribution function (cdf), say G , we define the generalized inverse by $G^{-1}(u) = \inf \{\tau \in \mathbb{R} : G(\tau) \geq u\}$ and write $G(\cdot-)$ to denote its left limit. Generally, for an arbitrary real-valued random variable X (e.g. stock return) with cdf F_X , the VaR at level $\alpha \in (0, 1)$, is given by $VaR_\alpha(X) = -F_X^{-1}(\alpha)$.² Let \mathcal{F}_{t-1} denote the σ -algebra generated by $\{\epsilon_u, u < t\}$. It follows that the conditional VaR of ϵ_t given \mathcal{F}_{t-1} at level $\alpha \in (0, 1)$ is

$$VaR_\alpha(\epsilon_t | \mathcal{F}_{t-1}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) VaR_\alpha(\eta_t). \quad (4.3)$$

For given α , the quantile of η_t is constant and can be treated as a parameter. Thus, denoting the cdf of η_t by F and setting $\xi_\alpha = F^{-1}(\alpha)$, equation (4.3) reduces to

$$VaR_\alpha(\epsilon_t | \mathcal{F}_{t-1}) = -\xi_\alpha \sigma_t(\theta_0). \quad (4.4)$$

Typically, α is fixed at a sufficiently small level such that $\xi_\alpha < 0$. Except for special cases (e.g. η_t possesses a standard normal distribution), ξ_α is unknown and needs to be estimated just like θ_0 .

4.3 Estimation

We estimate the parameters θ_0 and ξ_α following the two-step procedure of Francq and Zakoian (2015, Section 4.2). In the first step, we estimate the conditional volatility parameter θ_0 by Gaussian QML. This approach is motivated as follows: if the innovations $\{\eta_t\}$ were Gaussian, the variables $\eta_t(\theta) = \epsilon_t / \sigma_t(\theta)$ would be i.i.d. $N(0, 1)$ whenever $\theta = \theta_0$, where

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta). \quad (4.5)$$

The 'Q' in QML stands for 'quasi' and refers to the fact that F does not need to be the standard normal distribution function. Obviously, given a sample $\epsilon_1, \dots, \epsilon_n$, we generally cannot determine $\sigma_t(\theta)$ completely. Replacing the unknown presample observations by arbitrary values, say $\tilde{\epsilon}_t$, $t \leq 0$, we obtain

$$\tilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta), \quad (4.6)$$

²The negative sign is included to conform to the convention of reporting VaR as a positive number.

which serves as an approximation for (4.5). The QML estimator of θ_0 is defined by

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta) \quad (4.7)$$

with the criterion function specified by

$$\tilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\theta) \quad \text{and} \quad \tilde{\ell}_t(\theta) = -\frac{1}{2} \left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

In the second step, we estimate ξ_α on the basis of the first-step residuals, i.e. $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$. The empirical α -quantile of $\hat{\eta}_1, \dots, \hat{\eta}_n$ is given by

$$\hat{\xi}_{n,\alpha} = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t - z), \quad (4.8)$$

where $\rho_\alpha(u) = u(\alpha - \mathbb{1}_{\{u < 0\}})$ is the usual asymmetric absolute loss function (cf. Koenker and Xiao, 2006). Equivalently, we can write $\hat{\xi}_{n,\alpha} = \hat{\mathbb{F}}_n^{-1}(\alpha)$ with $\hat{\mathbb{F}}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}}$ being the empirical distribution function (edf) of the residuals.

Having obtained estimators for θ_0 and ξ_α , we turn to the estimation of the conditional VaR of the one-period ahead observation at level α . Whereas the notation $VaR_\alpha(\epsilon_{n+1} | \mathcal{F}_n)$ stresses the object's conditional nature, we henceforth proceed with the abbreviation $VaR_{n,\alpha}$ for notational convenience. Employing (4.6) – (4.8) we can estimate $VaR_{n,\alpha}$ by

$$\widehat{VaR}_{n,\alpha} = -\hat{\xi}_{n,\alpha} \tilde{\sigma}_{n+1}(\hat{\theta}_n). \quad (4.9)$$

Clearly, the estimator's large sample properties cannot be studied using traditional tools such as consistency since (4.9) does not permit a limit.

For the subsequent asymptotic analysis, we introduce the following assumptions.

Assumption 4.1. (*Compactness*) Θ is a compact subset of \mathbb{R}^r .

Assumption 4.2. (*Stationarity & ergodicity*) $\{\epsilon_t\}$ is a strictly stationary and ergodic solution of (4.1) with (4.2).

Assumption 4.3. (*Volatility process*) For any real sequence $\{x_i\}$, the function $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$ is continuous. Almost surely, $\sigma_t(\theta) > \underline{\omega}$ for any $\theta \in \Theta$ and

some $\underline{\omega} > 0$ and $\mathbb{E}[\sigma_t^s(\theta_0)] < \infty$ for some $s > 0$. Moreover, for any $\theta \in \Theta$, we assume $\sigma_t(\theta_0)/\sigma_t(\theta) = 1$ almost surely (a.s.) if and only if $\theta = \theta_0$.

Assumption 4.4. (*Initial conditions*) There exists a constant $\rho \in (0, 1)$ and a random variable C_1 measurable with respect to \mathcal{F}_0 and $\mathbb{E}[C_1^s] < \infty$ for some $s > 0$ such that

- (i) $\sup_{\theta \in \Theta} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \leq C_1 \rho^t$;
- (ii) $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives satisfying

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} \right\| \leq C_1 \rho^t, \quad \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho^t,$$

where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 4.5. (*Innovation process*) The innovations $\{\eta_t\}$ satisfy

- (i) $\eta_t \stackrel{iid}{\sim} F$ with F being continuous, $\mathbb{E}[\eta_t^2] = 1$ and η_t is independent of $\{\epsilon_u : u < t\}$;
- (ii) η_t admits a density f which is continuous and strictly positive around $\xi_\alpha < 0$;
- (iii) $\mathbb{E}[\eta_t^4] < \infty$.

Assumption 4.6. (*Interior*) θ_0 belongs to the interior of Θ denoted by $\mathring{\Theta}$.

Assumption 4.7. (*Non-degeneracy*) There does not exist a non-zero $\lambda \in \mathbb{R}^r$ such that $\lambda' \frac{\partial \sigma_t(\theta_0)}{\partial \theta} = 0$ a.s.

Assumption 4.8. (*Monotonicity*) For any real sequence $\{x_i\}$ and for any $\theta_1, \theta_2 \in \Theta$ satisfying $\theta_1 \leq \theta_2$ componentwise, we have $\sigma(x_1, x_2, \dots; \theta_1) \leq \sigma(x_1, x_2, \dots; \theta_2)$.

Assumption 4.9. (*Moments*) There exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that the following variables have finite expectation:

- (i) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^a,$
- (ii) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^b,$
- (iii) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^c$

for some a, b, c (to be specified).³

Assumption 4.10. (*Scaling stability*) There exists a function g such that for any $\theta \in \Theta$, for any $\lambda > 0$, and any real sequence $\{x_i\}$

$$\lambda\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; \theta_\lambda),$$

where $\theta_\lambda = g(\theta, \lambda)$ and g is differentiable in λ .

The previous set of assumptions is comparable to the conditions imposed by Francq and Zakoïan (2015). Regarding the innovation process we do not need to assume $\mathbb{E}[\eta_t] = 0$ (cf. Francq and Zakoïan, 2004, Remark 2.5). Whereas Cavaliere et al. (2018) assume the existence of the sixth moment of η_t for the fixed-design bootstrap in ARCH(q) models, we only require the fourth moment to be finite in Assumption 4.5(iii). In Assumption 4.8 the function $\sigma(x_1, x_2, \dots; \theta)$ is presumed to be monotonically increasing in θ , which is a feature shared by various stochastic volatility models (cf. Berkes and Horváth, 2003, Lemma 4.1). The monotonicity condition is used to establish the strong consistency of the quantile estimator. Further, we require higher order of moments in Assumption 4.9 for the bootstrap, which does not seem to be restrictive for the classical GARCH-type models (cf. Francq and Zakoïan, 2011, p. 165; Hamadeh and Zakoïan, 2011, p. 501). In particular, Assumption 4.9 is presumed to hold with $a = \pm 12$, $b = 12$ and $c = 6$ for establishing the convergence of the bootstrap information matrix (see Lemma 4.6 in Appendix 4.A.2).

On the basis of the previous assumptions we extend the strong consistency result of Francq and Zakoïan (2015, Theorem 1) to the quantile estimator.

Theorem 4.1. (*Strong consistency*) Under Assumptions 4.1–4.3, 4.4(i) and 4.5(i) the estimator in (4.7) is strongly consistent, i.e. $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$. If in addition Assumptions 4.6 and 4.9(i) hold with $a = -1$, then the estimator in (4.8) satisfies $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$.

Proof. Francq and Zakoïan (2015, Theorem 1) establish $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$. The second claim follows from $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$ (Lemma 4.1 in Appendix A.1) and van der Vaart (2000, Theorem 21.2). \square

³Note that the variables in (i) to (iii) are strictly stationary (Francq and Zakoïan, 2011, p. 181/406).

To lighten notation, we henceforth write $D_t(\theta) = \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}$ and drop the argument when evaluated at the true parameter, i.e. $D_t = D_t(\theta_0)$. The next result provides the joint asymptotic distribution of $\hat{\theta}_n$ and $\hat{\xi}_{n,\alpha}$ and is due to Francq and Zakoïan (2015).

Theorem 4.2. (*Asymptotic distribution*) Suppose Assumptions 4.1–4.7, 4.9 and 4.10 hold with $a = b = 4$ and $c = 2$. Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \sqrt{n}(\xi_\alpha - \hat{\xi}_{n,\alpha}) \end{pmatrix} \xrightarrow{d} N(0, \Sigma_\alpha) \quad \text{with} \quad \Sigma_\alpha = \begin{pmatrix} \frac{\kappa-1}{4} J^{-1} & \lambda_\alpha J^{-1} \Omega \\ \lambda_\alpha \Omega' J^{-1} & \zeta_\alpha \end{pmatrix}, \quad (4.10)$$

where $\kappa = \mathbb{E}[\eta_t^4]$, $\Omega = \mathbb{E}[D_t]$, $J = \mathbb{E}[D_t D_t']$, $\lambda_\alpha = \xi_\alpha \frac{\kappa-1}{4} + \frac{p_\alpha}{2f(\xi_\alpha)}$, $\zeta_\alpha = \xi_\alpha^2 \frac{\kappa-1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}$ and $p_\alpha = \mathbb{E}[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \alpha$.

Proof. See Francq and Zakoïan (2015, Theorem 4) and note that Assumption 4.10 is needed to ensure $\Omega' J^{-1} \Omega = 1$. \square

In a GARCH(p, q) setting Gao and Song (2008) quantify the uncertainty around $\hat{\theta}_n$ and $\hat{\xi}_{n,\alpha}$ using (4.10) while replacing the unknown quantities in Σ_α by estimates. In this spirit ξ_α can be substituted by $\hat{\xi}_{n,\alpha}$ and Ω, J, κ and p_α can be replaced by

$$\begin{aligned} \hat{\Omega}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t, & \hat{J}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t', \\ \hat{\kappa}_n &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^4, & \hat{p}_{n,\alpha} &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} - \alpha, \end{aligned} \quad (4.11)$$

respectively, with $\hat{D}_t = \tilde{D}_t(\hat{\theta}_n)$ and $\tilde{D}_t(\theta) = \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta}$. The strong consistency of the estimators in (4.11) follow from Theorem 4.1 and Lemma 4.2 in Appendix 4.A.1. Moreover, kernel smoothing is commonly employed to estimate the density f , i.e.

$$\hat{f}_n^S(x) = \frac{1}{nh_n} \sum_{t=1}^n k\left(\frac{x - \hat{\eta}_t}{h_n}\right) \quad (4.12)$$

with kernel function k and bandwidth $h_n > 0$. Whereas Gao and Song (2008) consider Lipschitz-continuous kernels, an alternative estimator is based on the uniform kernel $k(x) = \frac{1}{2} \mathbb{1}_{\{|x| \leq 1\}}$ yielding $\hat{f}_n^S(\hat{\xi}_{n,\alpha}) \xrightarrow{P} f(\xi_\alpha)$ whenever $h_n \sim n^{-e}$

for some $\varrho \in (0, 1/2]$.⁴ Based on (4.11) and (4.12), we obtain a consistent estimator for Σ_α denoted by $\hat{\Sigma}_{n,\alpha}$.

Employing Theorem 4.2 we can study the asymptotic behavior of the conditional VaR estimator in (4.9). Since the conditional volatility varies over time, a limiting distribution cannot exist and therefore the concept of weak convergence is not applicable in this context. In Chapter 2 we study a *merging* concept generalizing the notion of weak convergence, i.e. two sequences of (random) cdfs $\{F_n\}, \{G_n\}$ *merge* (in probability) if and only if their bounded Lipschitz distance $d_{BL}(F_n, G_n)$ converges to zero (in probability). Presuming two independent samples, one for parameter estimation and one for conditioning, the delta method suggests that the VaR estimator, centered at $VaR_{n,\alpha}$ and inflated by \sqrt{n} , and

$$N\left(0, \begin{pmatrix} -\xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta} \\ \sigma_{n+1} \end{pmatrix}' \Sigma_\alpha \begin{pmatrix} -\xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta} \\ \sigma_{n+1} \end{pmatrix}\right) \quad (4.13)$$

given \mathcal{F}_n merge in probability. Equation (4.13) highlights once more the relevance of the merging concept since its conditional variance still depends on n and does not converge as $n \rightarrow \infty$. Together with Theorem 4.1 and $\hat{\Sigma}_{n,\alpha} \xrightarrow{P} \Sigma_\alpha$, it yields a $100(1 - \gamma)\%$ confidence interval for $VaR_{n,\alpha}$ with bounds (cf. Francq and Zakoïan, 2015, Equation (23))

$$\widehat{VaR}_{n,\alpha} \pm \frac{\Phi^{-1}(\gamma/2)}{\sqrt{n}} \left\{ \begin{pmatrix} -\hat{\xi}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix}' \hat{\Sigma}_{n,\alpha} \begin{pmatrix} -\hat{\xi}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix} \right\}^{\frac{1}{2}}, \quad (4.14)$$

where Φ is the standard normal cdf. However, with the exception of perhaps some experimental settings, researchers rarely have a replicate, independent of the original series, at hand. An asymptotic justification for the interval on the basis of a single sample is given in Chapter 2. Nevertheless, the interval in (4.14) may perform rather poorly since the density estimation appears sensitive regarding the choice of bandwidth (see Gao and Song, 2008, Section 4). Bootstrap methods offer an alternative way to quantify the uncertainty around the estimators.

⁴It follows from Lemma 4.3 in Appendix 4.A.1, the mean value theorem and \sqrt{n} -consistency of $\hat{\xi}_{n,\alpha}$.

4.4 Bootstrap

Bootstrap approximations frequently provide better insight into the actual distribution than the asymptotic approximation, yet they require a careful setup. Hall and Yao (2003) show that conventional bootstrap methods are inconsistent in a GARCH model lacking finite fourth moment in the case of the squared innovations' distribution not being in the domain of attraction of the normal distribution. They consider a subsample bootstrap instead and study its asymptotic properties. In correspondence, an m -out-of- n without-replacement bootstrap is proposed by Spierdijk (2016) to construct confidence intervals for ARMA-GARCH VaR.

Pascual et al. (2006) present a residual bootstrap in a GARCH(1, 1) setting and assess its finite sample properties by means of simulation. Their bootstrap scheme follows a recursive design in which the bootstrap observations are generated iteratively using the estimated volatility dynamics. Building upon their results, Christoffersen and Gonçalves (2005) construct bootstrap confidence intervals for (conditional) VaR and Expected Shortfall and compare them to competitive methods within the GARCH(1, 1) model. Theoretical results on the recursive-design residual bootstrap are provided by Hidalgo and Zaffaroni (2007) and Jeong (2017) for the ARCH(∞) and GARCH(p, q) model, respectively.

In contrast, Shimizu (2010) considers fixed-design variants of the wild and the residual bootstrap in which the ARMA-GARCH dynamics of the bootstrap samples are kept fixed at the values of the original series. The bootstrap estimators are based on a single Newton-Raphson iteration simplifying the proofs of first-order asymptotic validity. Shimizu's approach for the residual bootstrap is also employed in a multivariate GARCH setting by Francq, Horváth, and Zakoïan (2016). Recently, Cavaliere et al. (2018) study the fixed-design residual bootstrap in the context of ARCH(q) models and propose a bootstrap Wald statistic based on a QML bootstrap estimator. While their theory has been developed independently to ours, their simulation study indicates that the fixed-design bootstrap performs as well as the recursive bootstrap.

4.4.1 Fixed-design Residual Bootstrap

We propose a fixed-design residual bootstrap procedure, described in Algorithm 4.1, to approximate the distribution of the estimators in (4.7) – (4.9).

Algorithm 4.1. (*Fixed-design residual bootstrap*)

1. For $t = 1, \dots, n$, generate $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$ and the bootstrap observation $\epsilon_t^* = \tilde{\sigma}_t(\hat{\theta}_n)\eta_t^*$, where $\tilde{\sigma}_t(\theta)$ and $\hat{\theta}_n$ are given in (4.6) and (4.7), respectively.
2. Calculate the bootstrap estimator

$$\hat{\theta}_n^* = \arg \max_{\theta \in \Theta} L_n^*(\theta) \quad (4.15)$$

with the bootstrap criterion function given by

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t^*(\theta) \quad \text{and} \quad \ell_t^*(\theta) = -\frac{1}{2} \left(\frac{\epsilon_t^*}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

3. For $t = 1, \dots, n$ compute the bootstrap residual $\hat{\eta}_t^* = \epsilon_t^* / \tilde{\sigma}_t(\hat{\theta}_n^*)$ and obtain

$$\hat{\xi}_{n,\alpha}^* = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z). \quad (4.16)$$

4. Obtain the bootstrap estimator of the conditional VaR

$$\widehat{VaR}_{n,\alpha}^* = -\hat{\xi}_{n,\alpha}^* \tilde{\sigma}_{n+1}(\hat{\theta}_n^*). \quad (4.17)$$

Remark 4.1. In contrast to the literature, the bootstrap errors are drawn with replacement from the residuals rather than the standardized residuals. In fact, re-centering would be inappropriate in the case of $\mathbb{E}[\eta_t] \neq 0$. In addition, re-scaling of the residuals is typically redundant as $\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 = 1$ is implied by $\hat{\theta}_n \in \mathring{\Theta}$ under Assumption 4.10; see Francq and Zakoïan, 2011, p. 182/406 and note that the solution requires $\hat{\theta}_n$ belonging to the interior (Francq and Zakoïan, Oct. 2018, personal communication).

Remark 4.2. The term ‘fixed-design’ refers to the fact that the bootstrap observations are generated using $\tilde{\sigma}_t(\hat{\theta}_n) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \hat{\theta}_n)$. In contrast, a recursive-design scheme replicates the model’s dynamic structure, i.e. $\epsilon_t^* = \sigma_t^* \eta_t^*$ with $\sigma_t^* = \sigma(\epsilon_{t-1}^*, \dots, \epsilon_1^*, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \hat{\theta}_n)$ and $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$, which is computationally more demanding. We refer to Appendix 4.B for a complete description. See also Cavaliere et al. (2018) for more theoretical insights on the difference in the design in an ARCH(q).

Remark 4.3. Whereas (4.15) involves a nonlinear optimization, Shimizu (2010) proposes a Newton-Raphson type bootstrap estimator instead. The Newton-Raphson bootstrap estimator corresponding to (4.15) is given by

$$\hat{\theta}_n^{*NR} = \hat{\theta}_n + \hat{J}_n^{-1} \frac{1}{2n} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1),$$

which can considerably speed up computations.

In the following subsection we show the asymptotic validity of the fixed-design bootstrap procedure described in Algorithm 4.1.

4.4.2 Bootstrap Consistency

Subsequently, we employ the usual notation for bootstrap asymptotics, i.e. “ $\xrightarrow{p^*}$ ” and “ $\xrightarrow{d^*}$ ”, as well as the standard bootstrap stochastic order symbol “ $o_{p^*}(1)$ ” (cf. Chang and Park, 2003). To prove the asymptotic validity of the proposed bootstrap procedure, we first focus on the stochastic volatility part. Since L_n^* is maximized at $\hat{\theta}_n^*$ its derivative is equal to zero: $\frac{\partial L_n^*(\hat{\theta}_n^*)}{\partial \theta} = 0$. A Taylor expansion around $\hat{\theta}_n$ yields

$$0 = \sqrt{n} \frac{\partial L_n^*(\hat{\theta}_n^*)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t^*(\hat{\theta}_n) + \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \right) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)$$

with $\check{\theta}_n$ lying between $\hat{\theta}_n^*$ and $\hat{\theta}_n$. Lemma 4.6 in Appendix 4.A.2 establishes $\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \xrightarrow{p^*} -2J$ almost surely. As $\frac{\partial}{\partial \theta} \ell_t^*(\theta) = \tilde{D}_t(\theta) \left(\frac{\epsilon_t^{*2}}{\tilde{\sigma}_t^2(\theta)} - 1 \right)$, the first term on the right hand side reduces to $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1)$. Hence, we get

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \frac{1}{2} J^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1) + o_{p^*}(1) \quad (4.18)$$

almost surely with $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1)$ converging in conditional distribution to $N(0, (\kappa - 1)J)$ almost surely by Lemma 4.7 in Appendix 4.A.2. The foregoing discussion can be summarized by the following intermediate result.

Proposition 4.1. *Suppose Assumptions 4.1–4.4, 4.5(i), 4.5(iii), 4.6, 4.7, 4.9 and*

4.10 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} N\left(0, \frac{\kappa - 1}{4} J^{-1}\right)$$

almost surely.

Proposition 4.1 establishes the asymptotic validity of the bootstrap for the conditional volatility parameters. Next, we turn to the estimator of the quantile parameter associated with the VaR at level α . Establishing the asymptotic validity of the bootstrap for the second part appears challenging since the bootstrap innovations are drawn from the discrete distribution $\hat{\mathbb{F}}_n$. To overcome this issue we rely on arguments employed by Bahadur (1966) and Berkes and Horváth (2003). Following the general steps of the proof of Francq and Zakoïan (2015, Theorem 4), we standardize equation (4.16) such that the bootstrap quantile estimator satisfies

$$\sqrt{n}(\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}) = \arg \min_{z \in \mathbb{R}} \underbrace{\sum_{t=1}^n \rho_\alpha\left(\hat{\eta}_t^* - \hat{\xi}_{n,\alpha} - \frac{z}{\sqrt{n}}\right) - \sum_{t=1}^n \rho_\alpha(\eta_t^* - \hat{\xi}_{n,\alpha})}_{Q_n^*(z)}.$$

Employing the identity of Koenker and Xiao (2006, Equation (A.3)) we obtain⁵

$$Q_n^*(z) = zX_n^* + Y_n^* + I_n^*(z) + J_n^*(z) \quad (4.19)$$

with

$$\begin{aligned} X_n^* &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha), \\ Y_n^* &= \sum_{t=1}^n (\eta_t^* - \hat{\eta}_t^*) (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha), \\ I_n^*(z) &= \sum_{t=1}^n \int_0^{\frac{z}{\sqrt{n}}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds, \\ J_n^*(z) &= \sum_{t=1}^n \int_{\frac{z}{\sqrt{n}}}^{\frac{z}{\sqrt{n}} + \eta_t^* - \hat{\eta}_t^*} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds. \end{aligned}$$

Subsequently, we look at each term in turn while resorting to Lemmas 4.7 to 4.10 in Appendix 4.A.2. Lemma 4.7 yields $X_n^* \xrightarrow{d^*} N(0, \alpha(1-\alpha))$ almost surely. Further,

⁵Note that the identity holds not only for $u \neq 0$ but also for $u = 0$.

we notice that Y_n^* neither depends on z nor interacts with it; therefore it can be disregarded. The term $I_n^*(z)$ converges in conditional probability to $\frac{z^2}{2}f(\xi_\alpha)$ in probability by Lemma 4.8. Next, we analyze the asymptotic properties of $J_n^*(z)$, which can be split into $J_n^*(z) = J_{n,1}^*(z) + J_{n,2}^*(z)$ with

$$J_{n,1}^*(z) = \sum_{t=1}^n \int_0^{\eta_t^* - \hat{\eta}_t^*} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} + s\}} - \mathbb{1}_{\{\eta_t^* - \hat{\xi}_{n,\alpha} - z/\sqrt{n} < 0\}}) ds \quad (4.20)$$

$$J_{n,2}^*(z) = \sum_{t=1}^n (\eta_t^* - \hat{\eta}_t^*) (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}). \quad (4.21)$$

Deviating from the proof of Francq and Zakoïan (2015), Lemma 4.9 shows that $J_{n,1}^*(z)$ converges in conditional distribution to a random variable, which does not depend on z , in probability. We refer to Remark 4.5 in Appendix A.2 for more technical details. Further, the second term is equal to $J_{n,2}^*(z) = z\xi_\alpha f(\xi_\alpha)\Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$ in probability by Lemma 4.10. By the preceding discussion we obtain

$$Q_n^*(z) = \frac{z^2}{2}f(\xi_\alpha) + z\left(X_n^* + \xi_\alpha f(\xi_\alpha)\Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\right) + J_{n,1}^*(z) + Y_n^* + o_{p^*}(1)$$

in probability. Employing Xiong and Li (2008, Theorem 3.3) and the basic corollary of Hjort and Pollard (2011), we obtain⁶

$$\sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) = \xi_\alpha \Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{f(\xi_\alpha)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) + o_{p^*}(1)$$

in probability. Together with (4.18) we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}J^{-1} & O_{r \times 1} \\ \frac{1}{2}\xi_\alpha \Omega' J^{-1} & \frac{1}{f(\xi_\alpha)} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) \end{pmatrix} + o_{p^*}(1).$$

Employing Lemma 4.7 leads to the chapter's main result.

Theorem 4.3. (*Bootstrap consistency*) *Suppose Assumptions 4.1–4.10 hold with*

⁶Matching notation, we take $A_n(z) = Q_n^*(z)$, which is convex, and set $B_n(z) = \frac{z^2}{2}V + zU_n + C_n$, where $V = f(\xi_\alpha)$, $U_n = X_n^* + \xi_\alpha f(\xi_\alpha)\Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ and $C_n + r_n(z) = J_{n,1}^*(z) + Y_n^* + o_{p^*}(1)$ with $r_n(z) \xrightarrow{p} 0$ for each $z \in \mathbb{R}$. The minimizers of $A_n(z)$ and $B_n(z)$ are $\alpha_n = \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*)$ and $\beta_n = -V^{-1}U_n$, respectively. The basic corollary of Hjort and Pollard (2011) states $\alpha_n - \beta_n = o_p(1)$, which implies $\alpha_n - \beta_n = o_{p^*}(1)$ in probability (Xiong and Li, 2008, Theorem 3.3).

$a = \pm 12$, $b = 12$ and $c = 6$. Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \end{pmatrix} \xrightarrow{d^*} N(0, \Sigma_\alpha)$$

in probability.

Theorem 4.3 is useful to validate the bootstrap for the conditional VaR estimator. For the asymptotic behavior of the conditional VaR estimator we refer to (4.13) and the text preceding it. The following corollary is established.

Corollary 4.1. *Under the assumptions of Theorem 4.3 the conditional cdfs of $\sqrt{n}(\widehat{VaR}_{n,\alpha}^* - \widehat{VaR}_{n,\alpha})$ given \mathcal{F}_n and (4.13) given \mathcal{F}_n merge in probability.*

Its proof is along the lines of the proof of Theorem 2.2. Having proven first-order asymptotic validity of the bootstrap procedure described in Section 4.4.1, we turn to constructing bootstrap confidence intervals for VaR.

4.4.3 Bootstrap Confidence Intervals for VaR

Clearly, the VaR evaluation in (4.9) is subject to estimation risk that needs to be quantified. We propose the following algorithm to obtain approximately $100(1 - \gamma)\%$ confidence intervals.

Algorithm 4.2. (*Fixed-design bootstrap confidence intervals for VaR*)

1. Acquire a set of B bootstrap replicates, i.e. $\widehat{VaR}_{n,\alpha}^{*(b)}$ for $b = 1, \dots, B$, by repeating Algorithm 4.1.
2. (i) Obtain the equal-tailed percentile (EP) interval

$$\left[\widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2), \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2) \right] \quad (4.22)$$

$$\text{with } \hat{G}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\left\{ \sqrt{n}(\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}) \leq x \right\}}.$$

- (ii) Calculate the reversed-tails (RT) interval

$$\left[\widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2) \right]. \quad (4.23)$$

(iii) Compute the symmetric (SY) interval

$$\left[\widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1-\gamma), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1-\gamma) \right] \quad (4.24)$$

$$\text{with } \hat{H}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}|\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}| \leq x\}}.$$

The interval in (4.22) is obtained by the EP method, that is frequently encountered in the bootstrap literature. “Flipping around” its tails leads to the RT interval given in (4.23), which can be motivated by the results of Falk and Kaufmann (1991).⁷ Clearly, the RT and the EP have equal length. Whereas (4.23) in its current form emphasizes the interval’s name, RT type intervals are frequently reported in their reduced form, i.e. the lower and upper bound of (4.23) simplify to the $\gamma/2$ and $1 - \gamma/2$ quantiles of $\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\widehat{VaR}_{n,\alpha}^{*(b)} \leq x\}}$, respectively. A RT type bootstrap interval for the VaR is also constructed in reduced form by Christoffersen and Gonçalves (2005). Last, the interval in (4.24) presumes symmetry for rationalizing its construction.

4.5 Numerical Illustration

4.5.1 Monte Carlo Experiment

In order to evaluate the finite sample performance of the proposed bootstrap procedure a Monte Carlo experiment is conducted. We confine ourselves to four conditional volatility specifications related to Examples 4.1 and 4.2 in Section 4.2. The first two are GARCH(1,1) parameterizations with

$$(i) \text{ high persistence: } \theta_0 = (0.05 \times 20^2 / 252, 0.15, 0.8)';$$

$$(ii) \text{ low persistence: } \theta_0 = (0.05 \times 20^2 / 252, 0.4, 0.55)';$$

which are similar to the specifications of Gao and Song (2008, Section 4) and Spierdijk (2016, Section 4.2). In addition, we study two T-GARCH(1,1) scenarios likewise associated with high and low persistence:

⁷In a random sample setting Falk and Kaufmann (1991) prove that the RT bootstrap interval for quantiles has asymptotically greater coverage than the corresponding EP bootstrap interval. For additional insights we refer to Hall and Martin (1988).

(iii) high persistence: $\theta_0 = (0.05 \times 20/\sqrt{252}, 0.05, 0.10, 0.8)'$;

(iv) low persistence: $\theta_0 = (0.05 \times 20/\sqrt{252}, 0.1, 0.3, 0.55)'$.

Within the experiment the VaR level takes two values, i.e. $\alpha \in \{0.01, 0.05\}$, and there are two possible innovation distributions: the standard normal distribution and a Student- t distribution with 6 degrees of freedom.⁸ We consider four estimation sample sizes, $n \in \{500; 1,000; 5,000; 10,000\}$, whereas the number of bootstrap replicates is fixed and equal to $B = 2,000$. For each model version we simulate $S = 2,000$ independent Monte Carlo trajectories.

All simulations are performed on a HP Z640 workstation with 16 cores using Matlab R2016a. The numerical optimization of the log-likelihood function is carried out employing the build-in function *fmincon* and running time is reduced by parallel computing using *parfor*.

Figure 4.1 displays the density of the distribution of the two-step QMLE estimator and the corresponding bootstrap distribution (given a particular sample) in the high persistence GARCH(1,1) case for $n = 5,000$. Figures 4.1(i) to 4.1(iii) indicate that the bootstrap distribution mimics adequately the finite sample distribution of the estimator of the volatility parameters. Besides, Figure 4.1(iv) illustrates that the bootstrap approximation works as well for the distribution of the quantile estimator. Moreover, all density plots are roughly bell-shaped supporting the theoretical implications of Theorem 4.2 and 4.3.

Table 4.1 reports the results of the three 90%-bootstrap intervals for the 5%-VaR when the innovation distribution is Student- t (henceforth referred to as baseline). In the GARCH(1,1) high persistence case (Panel I, right), we see that average coverage varies around 90% across all sample sizes for the RT and the SY interval. In contrast, the EP interval falls short of the nominal 90% by 5.75 percentage points (pp) for small sample size ($n = 500$). Nevertheless, its average coverage approaches the nominal value as the sample size increases. Remarkably, for all three intervals the average rate of the conditional VaR being below the interval is considerably less than the average rate of the conditional VaR being above the interval when the sample size is small ($n = 500$). Regarding the intervals' length, we observe that the SY interval is on average larger than the EP/RT interval. As the sample size increases this gap diminishes and the intervals' average lengths shrink. Considering the low persistent case (Panel I, left) we find

⁸The Student- t innovations are appropriately standardized to satisfy $\mathbb{E}\eta_t^2 = 1$.

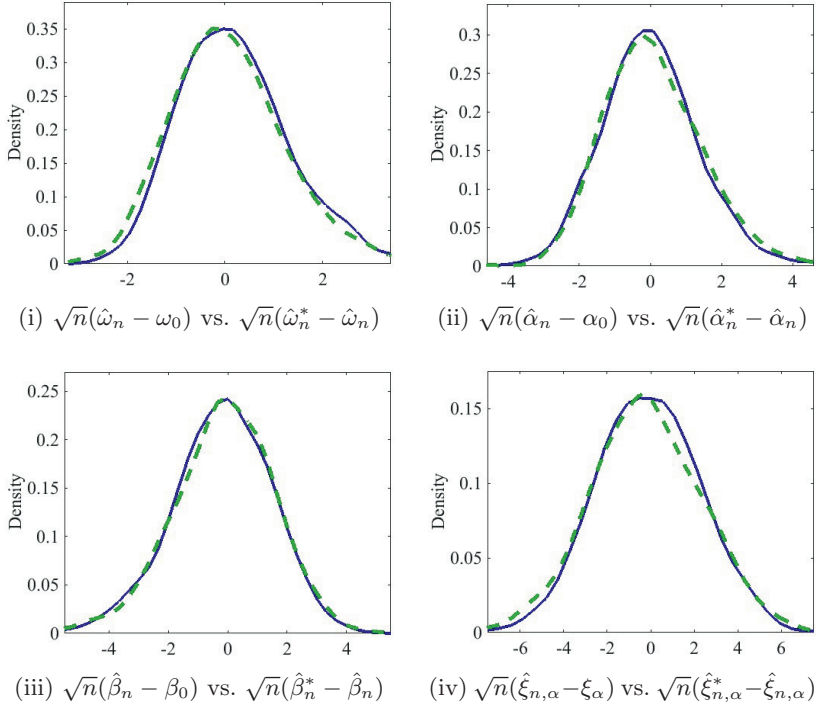


Figure 4.1: Density estimates for the distribution of the 2-step QMLE (full line) based on $S = 2,000$ simulations and the fixed-design bootstrap distribution (dashed line) based on $B = 2,000$ replications. α is set to 0.05 and the DGP is a GARCH(1, 1) with $\theta_0 = (0.08, 0.15, 0.8)'$, sample size $n = 5,000$ and (normalized) Student-t innovations (6 degrees of freedom).

similar results regarding the intervals' average coverage, yet their average lengths turn out to be smaller compared to the high persistent case. This is intuitive as the conditional volatility tends to vary less in the low persistent case. Regarding the T-GARCH(1, 1) in Panel II, the overall picture is similar as in the GARCH case, however the under-coverage of the EP interval in small samples appears to be more extreme.

Next, we consider deviations from the baseline specification. In particular, we study a change in the innovation distribution F (Table 4.2), a change in the VaR level α (Table 4.3) and a change in intervals' nominal coverage probability $100(1 - \gamma)\%$ (Table 4.4). While Table 4.5 draws attention to the average coverage gap between the EP and the RT bootstrap interval, Table 4.6 permits a comparison of the fixed-design bootstrap with its recursive-design counterpart.

4 A RESIDUAL BOOTSTRAP FOR CONDITIONAL VALUE-AT-RISK

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
Panel I: GARCH(1, 1)							
		low persistence			high persistence		
500	EP	84.50	6.30/9.20	0.431	84.25	6.30/9.45	0.582
	RT	91.50	3.75/4.75	0.431	91.45	3.40/5.15	0.582
	SY	90.40	3.60/6.00	0.443	90.10	3.65/6.25	0.596
1,000	EP	87.05	5.05/7.90	0.305	86.45	6.05/7.50	0.417
	RT	91.55	3.75/4.70	0.305	91.05	4.50/4.45	0.417
	SY	91.15	3.55/5.30	0.310	90.30	4.75/4.95	0.424
5,000	EP	87.45	6.15/6.40	0.144	87.85	5.70/6.45	0.191
	RT	90.35	5.30/4.35	0.144	89.50	5.25/5.25	0.191
	SY	89.75	5.35/4.90	0.145	89.70	4.80/5.50	0.192
10,000	EP	87.95	5.40/6.65	0.103	88.30	5.40/6.30	0.135
	RT	89.75	4.95/5.30	0.103	89.70	4.90/5.40	0.135
	SY	89.55	4.80/5.65	0.103	89.40	4.95/5.65	0.136
Panel II: T-GARCH(1, 1)							
		low persistence			high persistence		
500	EP	82.80	6.10/11.10	0.104	82.35	6.25/11.40	0.214
	RT	90.20	4.20/5.60	0.104	91.30	3.50/5.20	0.214
	SY	89.15	4.05/6.80	0.107	90.10	2.95/6.95	0.219
1,000	EP	84.45	6.00/9.55	0.076	82.95	6.90/10.15	0.156
	RT	90.10	4.60/5.30	0.076	90.75	4.50/4.75	0.156
	SY	89.00	4.35/6.65	0.077	89.10	4.55/6.35	0.159
5,000	EP	88.40	5.35/6.25	0.035	88.30	4.95/6.75	0.073
	RT	90.35	5.20/4.45	0.035	90.45	4.80/4.75	0.073
	SY	90.75	4.70/4.55	0.035	89.75	4.45/5.80	0.074
10,000	EP	87.85	4.95/7.20	0.026	87.95	5.40/6.65	0.053
	RT	89.75	4.60/5.65	0.026	89.55	5.10/5.35	0.053
	SY	89.35	4.45/6.20	0.026	89.30	4.95/5.75	0.054

Table 4.1: The table reports distinct features of the **fixed-design** bootstrap confidence intervals for the conditional VaR at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes (n), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 degrees of freedom)**, whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
Panel I: GARCH(1, 1)							
		low persistence			high persistence		
500	EP	85.10	6.75/8.15	0.384	83.10	7.75/9.15	0.472
	RT	91.45	3.10/5.45	0.384	89.70	3.60/6.70	0.472
	SY	90.85	3.50/5.65	0.396	88.65	4.20/7.15	0.482
1,000	EP	85.25	7.10/7.65	0.261	87.55	5.55/6.90	0.335
	RT	91.00	3.50/5.50	0.261	91.10	3.25/5.65	0.335
	SY	89.50	4.30/6.20	0.266	90.85	3.55/5.60	0.340
5,000	EP	87.50	5.30/7.20	0.121	87.85	5.55/6.60	0.149
	RT	90.20	4.35/5.45	0.121	89.30	4.85/5.85	0.149
	SY	89.75	4.30/5.95	0.122	89.15	4.95/5.90	0.150
10,000	EP	88.85	5.65/5.50	0.086	89.25	5.30/5.45	0.105
	RT	90.50	4.55/4.95	0.086	90.30	4.70/5.00	0.105
	SY	90.40	4.85/4.75	0.087	90.10	4.95/4.95	0.106
Panel II: T-GARCH(1, 1)							
		low persistence			high persistence		
500	EP	85.15	5.90/8.95	0.086	83.50	6.65/9.85	0.173
	RT	90.10	3.30/6.60	0.086	90.20	2.85/6.95	0.173
	SY	89.45	3.75/6.80	0.088	89.15	3.60/7.25	0.178
1,000	EP	84.80	5.95/9.25	0.061	84.60	6.60/8.80	0.125
	RT	90.05	3.85/6.10	0.061	90.90	3.25/5.85	0.125
	SY	89.50	3.85/6.65	0.062	89.55	4.05/6.40	0.128
5,000	EP	87.95	5.30/6.75	0.028	86.85	5.60/7.55	0.057
	RT	89.90	4.40/5.70	0.028	88.65	4.50/6.85	0.057
	SY	89.55	4.55/5.90	0.028	88.35	4.65/7.00	0.058
10,000	EP	89.70	4.70/5.60	0.020	88.60	5.25/6.15	0.041
	RT	90.60	4.40/5.00	0.020	90.50	4.45/5.05	0.041
	SY	90.95	4.15/4.90	0.020	90.25	4.65/5.10	0.041

Table 4.2: The table reports distinct features of the **fixed-design** bootstrap confidence intervals for the conditional VaR at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes (n), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with **Gaussian innovations**, whereas in Panel II the DGP is a Gaussian T-GARCH(1,1).

The simulation results for the scenario when the η_t 's follow a standard normal distribution are tabulated in Table 4.2. Although the error distribution underlying the QMLE is correctly specified in this case, the qualitative results stated above with regard to Table 4.1 persist: the RT and the SY intervals possess accurate coverage rates across sample sizes, whereas the EP interval exhibits under-coverage in samples of modest size. Although falling short of the nominal coverage value, the EP intervals exhibit a tendency of improved average coverage, e.g. 83.50% in the high-persistent T-GARCH case with $n = 500$ compared to 80.45% in Table 4.1. Moreover, we observe that the intervals are on average shorter in the Gaussian case than in the baseline case. This seems partially driven by a smaller variance of $\hat{\xi}_{n,\alpha}$; for $\alpha = 0.05$ the asymptotic variance ζ_α in (4.10) is equal to 3.11 in the Gaussian case compared to 5.72 in the Student-t case with 6 degrees of freedom.

Table 4.3 focuses on the VaR at level $\alpha = 0.01$ instead. In comparison to Table 4.1 it is striking that the EP interval performs worse in terms of average coverage (especially for smaller sample sizes). Take note that this attribute is mainly driven by differences in the right tail of the bootstrap density. In contrast, the average coverage of the RT and the SY interval remain varying around 90% for $n \geq 5,000$ while a small loss of accuracy occurs in shorter samples. Coherent with a value of ζ_α around 32 at $\alpha = 0.01$ in the Student-t case, we find the intervals for the 1%-VaR to be on average considerably longer than the intervals for the 5%-VaR in the baseline case.

Increasing the intervals' nominal value from 90% to 95%, Table 4.4 presents the results of the three 95%-bootstrap intervals for the 5%-VaR. Again, the RT and the SY intervals perform well in terms of coverage: across sample sizes their average coverages are fairly close to 95%. Once more, the EP interval falls short of the nominal coverage value, yet the discrepancy appears to be less in comparison to the baseline. For example in the high-persistent GARCH case with $n = 500$, the EP interval falls short by $95\% - 90.25\% = 4.75\text{pp}$ compared to $90\% - 84.25\% = 5.75\text{pp}$ (see Table 4.1).

The question arises why the EP interval performs worse than the other two interval types in small samples, which seems counter-intuitive at first. Howbeit the results are in line with the theoretical findings of Falk and Kaufmann (1991, unnumbered Corollary, p. 488). In a random sample setting they prove that the RT bootstrap interval for quantiles has asymptotically greater coverage than the

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
Panel I: GARCH(1, 1)							
		low persistence			high persistence		
500	EP	78.40	7.40/14.20	0.918	79.65	7.00/13.35	1.227
	RT	89.45	2.40/8.15	0.918	89.70	2.05/8.25	1.227
	SY	87.85	2.60/9.55	0.955	88.55	2.60/8.85	1.272
1,000	EP	81.45	5.75/12.80	0.657	82.00	5.60/12.40	0.886
	RT	90.40	2.30/7.30	0.657	89.90	3.05/7.05	0.886
	SY	88.95	2.85/8.20	0.679	88.80	3.20/8.00	0.914
5,000	EP	85.30	5.80/8.90	0.306	85.95	5.05/9.00	0.407
	RT	91.30	3.60/5.10	0.306	91.05	3.50/5.45	0.407
	SY	90.45	3.65/5.90	0.312	90.40	3.40/6.20	0.413
10,000	EP	86.25	6.30/7.45	0.218	87.00	5.75/7.25	0.289
	RT	90.45	4.80/4.75	0.218	90.20	4.75/5.05	0.289
	SY	89.35	5.30/5.35	0.221	89.45	4.90/5.65	0.292
Panel II: T-GARCH(1, 1)							
		low persistence			high persistence		
500	EP	77.95	7.00/15.05	0.219	77.70	7.70/14.60	0.449
	RT	88.35	2.20/9.45	0.219	88.65	1.70/9.65	0.449
	SY	86.65	2.60/10.75	0.228	88.10	1.95/9.95	0.467
1,000	EP	80.55	5.50/13.95	0.160	79.60	6.55/13.85	0.330
	RT	89.95	2.10/7.95	0.160	89.45	2.55/8.00	0.330
	SY	87.75	2.60/9.65	0.165	87.25	3.20/9.55	0.341
5,000	EP	86.25	4.85/8.90	0.074	85.50	5.55/8.95	0.155
	RT	91.40	3.70/4.90	0.074	91.80	3.70/4.50	0.155
	SY	90.20	3.60/6.20	0.075	90.25	3.75/6.00	0.157
10,000	EP	87.50	5.25/7.25	0.054	86.50	6.10/7.40	0.112
	RT	91.25	4.00/4.75	0.054	90.90	4.50/4.60	0.112
	SY	90.25	4.05/5.70	0.055	89.85	4.65/5.50	0.114

Table 4.3: The table reports distinct features of the **fixed-design** bootstrap confidence intervals for the conditional VaR at **level $\alpha = 0.01$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes (n), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 degrees of freedom)**, whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

4 A RESIDUAL BOOTSTRAP FOR CONDITIONAL VALUE-AT-RISK

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
Panel I: GARCH(1, 1)							
		low persistence			high persistence		
500	EP	90.20	3.25/6.55	0.515	90.25	3.30/6.45	0.696
	RT	96.00	1.70/2.30	0.515	96.40	1.45/2.15	0.696
	SY	95.55	1.35/3.10	0.534	95.15	1.50/3.35	0.720
1,000	EP	92.65	2.45/4.90	0.364	91.80	3.45/4.75	0.498
	RT	96.10	2.05/1.85	0.364	95.65	2.20/2.15	0.498
	SY	95.75	1.40/2.85	0.373	95.30	2.00/2.70	0.510
5,000	EP	92.95	3.45/3.60	0.171	93.25	2.85/3.90	0.228
	RT	95.65	2.15/2.20	0.171	95.30	2.20/2.50	0.228
	SY	94.90	2.50/2.60	0.173	95.05	2.20/2.75	0.230
10,000	EP	93.05	3.00/3.95	0.122	93.20	2.90/3.90	0.161
	RT	94.70	2.65/2.65	0.122	94.70	2.45/2.85	0.161
	SY	94.45	2.65/2.90	0.123	94.10	2.35/3.55	0.162
Panel II: T-GARCH(1, 1)							
		low persistence			high persistence		
500	EP	88.70	3.50/7.80	0.125	88.45	3.75/7.80	0.256
	RT	95.60	1.90/2.50	0.125	96.25	1.30/2.45	0.256
	SY	94.40	1.60/4.00	0.129	94.85	1.45/3.70	0.266
1,000	EP	89.90	3.65/6.45	0.090	90.50	3.40/6.10	0.186
	RT	95.55	2.00/2.45	0.090	95.45	1.85/2.70	0.186
	SY	94.70	2.00/3.30	0.093	94.50	1.95/3.55	0.192
5,000	EP	93.70	2.65/3.65	0.042	93.55	2.30/4.15	0.087
	RT	95.50	2.50/2.00	0.042	95.65	2.40/1.95	0.087
	SY	95.20	2.30/2.50	0.042	95.45	2.00/2.55	0.088
10,000	EP	93.90	2.30/3.80	0.031	93.25	2.90/3.85	0.064
	RT	95.10	2.50/2.40	0.031	94.95	2.35/2.70	0.064
	SY	94.95	2.25/2.80	0.031	94.70	2.35/2.95	0.064

Table 4.4: The table reports distinct features of the **fixed-design** bootstrap confidence intervals for the conditional VaR at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 95\%$** . For each interval type and different sample sizes (n), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 degrees of freedom)**, whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

corresponding EP bootstrap interval. The emerging gap⁹

- (i) tends to be smaller for larger sample sizes,
- (ii) tends to be larger for more extreme quantiles and
- (iii) tends to vary with the nominal coverage rate in a non-monotonic way.

Sample size	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel I: GARCH(1,1)								
	low persistence				high persistence			
500	7.00	6.35	11.05	5.80	7.20	6.60	10.05	6.15
1,000	4.50	5.75	8.95	3.45	4.60	3.55	7.90	3.85
5,000	2.90	2.70	6.00	2.70	1.65	1.45	5.10	2.05
10,000	1.80	1.65	4.20	1.65	1.40	1.05	3.20	1.50
Panel II: T-GARCH(1,1)								
	low persistence				high persistence			
500	7.40	4.95	10.40	6.90	8.95	6.70	10.95	7.80
1,000	5.65	5.25	9.40	5.65	7.80	6.30	9.85	4.95
5,000	1.95	1.95	5.15	1.80	2.15	1.80	6.30	2.10
10,000	1.90	0.90	3.75	1.20	1.60	1.90	4.40	1.70

Table 4.5: The table reports the average coverage gap between the RT and the EP fixed-design bootstrap interval in percentage points. For different sample sizes (n) Panel I presents the results for the low and high persistence parameterization of a GARCH(1,1), whereas Panel II displays the results for the corresponding T-GARCH(1,1) processes.

- (1) - Table 4.1: 5%-VaR, Student-t innovations and 90% nominal coverage (baseline)
- (2) - Table 4.2: 5%-VaR, Gaussian innovations and 90% nominal coverage
- (3) - Table 4.3: 1%-VaR, Student-t innovations and 90% nominal coverage
- (4) - Table 4.4: 5%-VaR, Student-t innovations and 95% nominal coverage

Table 4.5 presents the average coverage gap between the EP and the RT bootstrap interval of the conditional VaR. For example, in the low persistence GARCH(1,1) case of the baseline with $n = 500$, the average coverage gap amounts to $91.50\% - 84.50\% = 7.00\text{pp}$ (see also Table 4.1). It is striking that all values are

⁹We neglect their $o(n^{-1/2})$ term. Take note that the theoretical results of Falk and Kaufmann (1991) are not directly applicable in our setting due to GARCH-type effects.

positive within Table 4.5, which highlights the superiority of the RT bootstrap interval over the EP bootstrap interval. Further, it is eminent that average coverage gap tends to decrease with increasing sample size, which supports (i). Comparing columns (1) and (3) we also find that the average coverage gap tends to be larger for the 1%–VaR than for the 5%–VaR, which gives rise to (ii). Regarding (iii), the result of Falk and Kaufmann (1991, unnumbered Corollary, p. 488) suggests that the gap slightly decreases when increasing the nominal coverage from 90% to 95%. Such tendency is precisely observed when comparing columns (1) and (4) of Table 4.5.

With regard to Remark 4.2 in Section 4.4, Table 4.6 reports the simulation results for the recursive-design bootstrap. We refer to Appendix 4.B for computational details. In comparison to the fixed-design approach (see Table 4.1) we find that the recursive-design method performs similarly in terms of average coverage for each interval type, which corresponds to the simulation results of Cavaliere et al. (2018). It is striking, however, that the intervals’ average lengths are larger in the recursive-design than in the fixed-design setup. For example, in the high persistence GARCH case (Panel I, right) for $n = 500$ the average length in the recursive-design approach is 0.605 for the EP/RT interval compared to 0.582 in the fixed-design. As the sample size increases this difference disappears. Regarding the running time, the fixed-design bootstrap scheme operates faster than its recursive-design counterpart, e.g. in the T-GARCH high persistence case for $n = 500$, applying Algorithm 4.2 with $B = 2,000$ takes roughly 2.7 seconds whereas its recursive-design competitor takes about 2.9 seconds per simulation.

In summary, the simulations suggest that for both bootstrap designs the RT and the SY bootstrap interval work well in terms of average coverage even though their tails are unequally represented for smaller sample sizes. In contrast, for both bootstrap designs the EP interval falls short of its nominal coverage, which is in line with the theoretical findings of Falk and Kaufmann (1991). Since the fixed RT method leads on average to shorter intervals than the corresponding SY method and its recursive-design counterpart, this suggests to favor the fixed-design RT bootstrap interval in (4.23).

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
Panel I: GARCH(1, 1)							
		low persistence			high persistence		
500	EP	85.00	5.95/9.05	0.442	85.05	5.45/9.50	0.605
	RT	91.05	4.20/4.75	0.442	91.25	3.95/4.80	0.605
	SY	91.40	3.15/5.45	0.459	91.05	3.05/5.90	0.629
1,000	EP	87.00	4.50/8.50	0.309	86.50	5.55/7.95	0.425
	RT	91.60	4.00/4.40	0.309	91.20	4.45/4.35	0.425
	SY	91.70	3.15/5.15	0.317	91.00	4.05/4.95	0.436
5,000	EP	87.75	6.25/6.00	0.144	87.90	5.50/6.60	0.191
	RT	90.10	5.20/4.70	0.144	89.80	5.15/5.05	0.191
	SY	90.05	5.10/4.85	0.146	89.70	4.80/5.50	0.193
10,000	EP	87.95	5.45/6.60	0.103	89.00	4.80/6.20	0.135
	RT	89.70	5.00/5.30	0.103	89.30	5.35/5.35	0.135
	SY	89.50	4.95/5.55	0.103	89.15	5.10/5.75	0.136
Panel II: T-GARCH(1, 1)							
		low persistence			high persistence		
500	EP	82.90	5.65/11.45	0.106	82.65	6.00/11.35	0.216
	RT	89.80	4.60/5.60	0.106	91.45	3.50/5.05	0.216
	SY	89.50	3.90/6.60	0.110	90.50	2.90/6.60	0.224
1,000	EP	84.50	6.00/9.50	0.077	83.25	6.80/9.95	0.158
	RT	90.30	4.70/5.00	0.077	90.55	4.55/4.90	0.158
	SY	89.90	3.95/6.15	0.079	89.70	4.20/6.10	0.162
5,000	EP	88.15	5.45/6.40	0.035	88.40	4.80/6.80	0.074
	RT	90.25	5.35/4.40	0.035	90.10	5.15/4.75	0.074
	SY	90.50	4.90/4.60	0.036	90.50	4.15/5.35	0.075
10,000	EP	87.80	5.00/7.20	0.026	88.50	5.10/6.40	0.054
	RT	89.35	4.90/5.75	0.026	89.80	5.05/5.15	0.054
	SY	89.50	4.40/6.10	0.026	89.75	4.65/5.60	0.054

Table 4.6: The table reports distinct features of the **recursive-design** bootstrap confidence intervals for the conditional VaR at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes (n), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 degrees of freedom)**, whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

4.5.2 Empirical Application

We analyze the French stock market index CAC 40 for the period January 1, 1998 – July 1, 2018. The index values for the period are retrieved from Yahoo Finance and daily (log-) returns (expressed in %) are computed using $\epsilon_t = 100 \log(p_t/p_{t-1})$, where p_t denotes the closing value of the index at trading day t . Figure 4.2(i) dis-

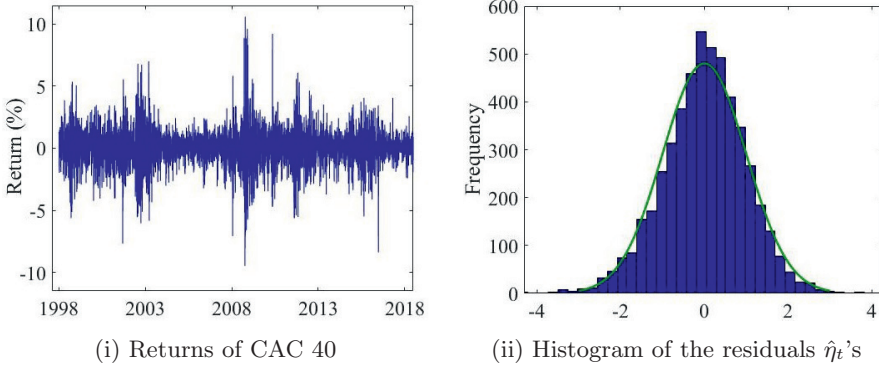


Figure 4.2: The returns of the French stock market index CAC 40 are plotted in (i) for the period January 1, 1998 – July 1, 2018. The histogram of the residuals is plotted in (ii) after fitting a T-GARCH(1,1) model to the subperiod January 1, 1998 – December 31, 2017. A scaled normal density is superimposed.

plays the resulting series of returns. We disregard the observations of the year 2018, which we leave for the out-of-sample evaluation, yielding $n = 5,100$ remaining observations (i.e. Jan. 1, 1998 - Dec. 31, 2017). For the volatility process we consider the T-GARCH(1,1) model specified in Example 4.2.¹⁰ Table 4.7 reports the corresponding point estimates with standard errors obtained by bootstrapping based on Algorithm 4.1. As documented in numerous studies we find that the volatility persistence is close to unity. In contrast, the point estimate $\hat{\alpha}_n^+$ is rather small. Further, we observe that $\hat{\alpha}_n^-$ is considerably larger than $\hat{\alpha}_n^+$ indicating a strong leverage effect, i.e. negative returns tend to increase volatility by more than positive returns of the same magnitude. Figure 4.2(ii) plots the histogram of the residuals with the normal distribution superimposed. We find that a (normalized) Student-t distribution with 10 degrees of freedom provides an improved

¹⁰We also consider an Asymmetric Power GARCH model (Ding, Granger, and Engle, 1993), i.e. $\sigma_{t+1}^\delta = \omega_0 + \alpha_0^+(\epsilon_t^+)^{\delta} + \alpha_0^-(\epsilon_t^-)^{\delta} + \beta_0\sigma_t^\delta$ with $\delta > 0$, which nests the GARCH(1,1) model ($\delta = 2$, $\alpha_0^+ = \alpha_0^-$) and the T-GARCH(1,1) model ($\delta = 1$) of Examples 4.1 and 4.2. In practice, the impact of the power δ on the volatility is minor and the QML approach of Hamadeh and Zakoian (2011) suggests a δ of 1.03 in favor for the T-GARCH specification.

	$\hat{\omega}_n$	$\hat{\alpha}_n^+$	$\hat{\alpha}_n^-$	$\hat{\beta}_n$
point estimate	0.0246	0.0150	0.1340	0.9237
std. error	0.0039	0.0099	0.0112	0.0084

Table 4.7: T-GARCH(1,1) estimates for the subperiod January 1, 1998 – December 31, 2017. The standard errors are obtained by applying the fixed-design residual bootstrap with $B = 2,000$ bootstrap replications.

fit, for which Assumption 4.5(iii) is met.

Next, we perform a rolling window analysis starting with subperiod January 1, 1998 – December 31, 2017 and ending with subperiod July 8, 1998 – June 30, 2018. We have 125 subperiods each consisting of $n = 5,100$ observations. For each rolling window period we fit a T-GARCH(1,1) model and estimate the one-period-ahead conditional VaR associated with level $\alpha = 0.05$. Further, we obtain the associated 95%-confidence intervals based on bootstrap (fixed-design RT and recursive-design RT) and asymptotic normality, where the latter is given in (4.14). For example, for the first window the T-GARCH(1,1) estimates are reported in Table 4.7 and the conditional 5%-VaR of the one-period ahead (i.e. January 2, 2018) is estimated by 1.48. The corresponding intervals are $[1.39, 1.58]$ (fixed-design), $[1.38, 1.57]$ (recursive-design) and $[1.41, 1.55]$ (asymptotic normality). Whereas both bootstrap methods yield RT intervals of (approx.) equal length, it is striking that the interval based on the asymptotic approximation is considerably shorter. The results of the rolling window analysis are visualized in Figure 4.3. It plots the realized return together with (the opposite of) the estimated conditional VaR. For clarity we only indicate the lower and upper bound of the 95% RT fixed-design bootstrap interval. We observe that in more turbulent times (e.g. February, 2018), the estimated VaR amplifies. In such volatile periods we expect the estimation risk to increase and, accordingly, we find wider bootstrap confidence intervals.

4.6 Concluding Remarks

We study the two-step estimation procedure of Francq and Zakoïan (2015) associated with the conditional VaR. In the first step, the conditional volatility parameters are estimated by QMLE, while the second step corresponds to approximating the quantile of the innovations' distribution by the empirical quantile of the resid-

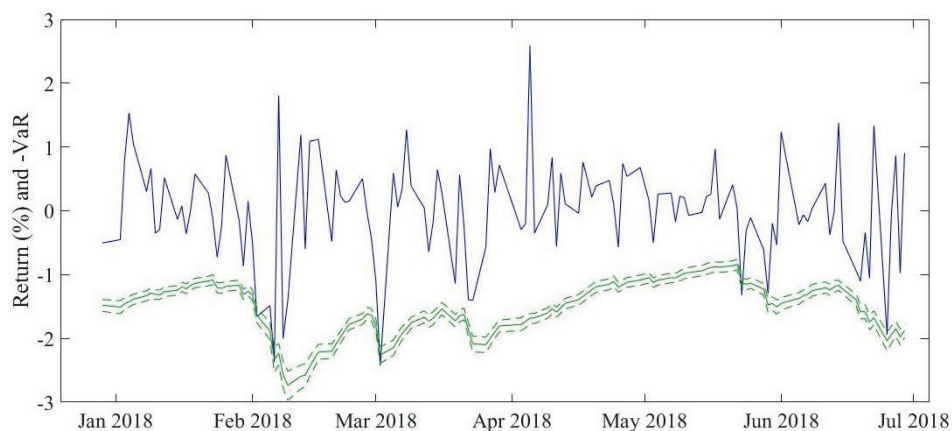


Figure 4.3: Returns (solid, blue) and the estimated conditional VaR (solid, green) for the period January 1, 2018 – July 1, 2018. The estimation rests on the 5,100 preceding observations. Lower and upper bounds for the conditional VaR (dashed, green) are based on the fixed-design bootstrap scheme using the RT method with $1 - \gamma = 95\%$.

uals. A fixed-design residual bootstrap method is proposed to mimic the finite sample distribution of the two-step estimator and its consistency is proven under mild assumptions. In addition, an algorithm is provided for the construction of bootstrap intervals for the conditional VaR to take into account the uncertainty induced by estimation. Three interval types are suggested and a large-scale simulation study is conducted to investigate their performance in finite samples. We find that the equal-tailed percentile interval based on the fixed-design residual bootstrap tends to fall short of its nominal value, whereas the corresponding interval based on reversed tails yields accurate average coverage combined with the shortest average length. Although the result seems counter-intuitive at first, it is in line with the theoretical findings of Falk and Kaufmann (1991). In the simulation study we also consider the recursive-design residual bootstrap. It turns out that the recursive-design and the fixed-design bootstrap perform similar in terms of average coverage. Yet in smaller samples the fixed-design scheme leads on average to shorter intervals. Further, the interval estimation by means of the fixed-design residual bootstrap is illustrated in an empirical application to daily returns of the French stock index CAC 40.

4.A Auxiliary Results and Proofs

4.A.1 Non-bootstrap Lemmas

In analogy to $D_t(\theta)$ and \hat{D}_t we write $H_t(\theta) = \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'}$ and $\hat{H}_t = \tilde{H}_t(\hat{\theta}_n)$ with $\tilde{H}_t(\theta) = \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'}$. Further, we introduce

$$\begin{aligned} S_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}, & T_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}, \\ U_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \|D_t(\theta)\|, & V_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \|H_t(\theta)\|, \end{aligned} \quad (4.25)$$

and stress that $\{S_t\}$, $\{T_t\}$, $\{U_t\}$ and $\{V_t\}$ are strictly stationary and ergodic processes (cf. Francq and Zakoian, 2011, p. 182/405).

Lemma 4.1. *Suppose Assumptions 4.1, 4.2, 4.3, 4.4(i), 4.5(i), 4.6 and 4.9(i) hold with $a = -1$. Then, we have $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$.*

Proof. The proof follows Berkes and Horváth (2003, Theorem 2.1 and Lemma 5.1) and consists of three parts. First, we show that for any $\varepsilon > 0$ there is a $\tau > 0$ such that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} - F(x) \right| \\ \leq 2 \left(F(x + \varepsilon |x|) - F(x - \varepsilon |x|) \right) \end{aligned} \quad (4.26)$$

almost surely for any $x \in \mathbb{R}$, where $\mathcal{V}_\tau(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \tau\}$. In the second step, we show $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$ for any $x \in \mathbb{R}$ using (4.26) and thereafter prove $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$.

Let $\varepsilon > 0$ and note that $\sigma_t \geq \underline{\omega}$ by Assumption 4.3. Together with Assumption 4.4(i), there exists a random variable n_0 such that $C_1 \rho^t / \sigma_t(\theta_0) \leq \varepsilon$ for all $t > n_0$. Then

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \sigma_t(\theta) / \sigma_t(\theta_0) + |x| C_1 \rho^t / \sigma_t(\theta_0)\}} \\ &\leq \frac{n_0}{n} + \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \sigma_t(\theta) / \sigma_t(\theta_0) + \varepsilon |x|\}} \end{aligned}$$

holds almost surely. Let $\tau > 0$ (to be specified); for any $\theta \in \mathcal{V}_\tau(\theta_0)$ we get

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} \leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}}$$

almost surely. The uniform ergodic theorem for strictly stationary sequences (cf. Francq and Zakoian, 2011, p. 181), henceforth called the uniform ergodic theorem, and Assumptions 4.2, 4.3 and 4.5(i) yield

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} &\xrightarrow{a.s.} \mathbb{E} \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} \\ &= \mathbb{E} F\left(\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\right). \end{aligned}$$

Assumptions 4.3 and 4.9(i) with $a = -1$ imply $\lim_{\tau \rightarrow 0} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) = x$ almost surely. Thus, the dominated convergence theorem entails

$$\lim_{\tau \rightarrow 0} \mathbb{E} F\left(\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\right) = F(x + \varepsilon|x|).$$

Putting the results together, we get that for every $\varepsilon > 0$, there is a $\tau > 0$ such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\tilde{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} \leq F(x) + 2\left(F(x + \varepsilon|x|) - F(x)\right)$$

almost surely for any $x \in \mathbb{R}$. Similarly it can be shown that for every $\varepsilon > 0$, there is a $\tau > 0$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\tilde{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} \geq F(x) - 2\left(F(x) - F(x - \varepsilon|x|)\right).$$

almost surely for any $x \in \mathbb{R}$. Combining both results, we establish (4.26).

Next, we show $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$ for any $x \in \mathbb{R}$. Let $\delta > 0$; by continuity of F (see Assumption 4.5(i)), there is a $\varepsilon > 0$ such that $|F(x + \varepsilon|x|) - F(x - \varepsilon|x|)| < \delta/2$. Employing equation (4.26), there are $\tau > 0$ and a random variable n_1 such that

$$\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\tilde{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} - F(x) \right| < \delta$$

for all $n \geq n_1$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ by Theorem 4.1 there is a random variable n_2 such that $\hat{\theta}_n \in \mathcal{V}_\tau(\theta_0)$ for all $n \geq n_2$. Thus,

$$|\hat{\mathbb{F}}_n(x) - F(x)| \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} - F(x) \right| < \delta$$

for all $n \geq \max\{n_1, n_2\}$, which establishes $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$ for any $x \in \mathbb{R}$. Using Pólya's lemma (cf. Roussas, 1997, p. 206), we establish $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$ completing the proof. \square

Lemma 4.2. *Suppose Assumptions 4.1–4.3, 4.4(i) and 4.5(i) hold.*

- (i) *If in addition Assumptions 4.4(ii) and 4.9(ii) hold with $b = 1$, then $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$.*
- (ii) *If in addition Assumptions 4.4(ii) and 4.9(ii) hold with $b = 2$, then $\hat{J}_n \xrightarrow{a.s.} J$.*
- (iii) *If in addition Assumptions 4.4(ii) and 4.9(iii) hold with $c = 1$, then $\frac{1}{n} \sum_{t=1}^n \hat{H}_t \xrightarrow{a.s.} \mathbb{E}[H_t]$.*
- (iv) *If in addition Assumptions 4.5(iii) and 4.9(i) hold with $a = 4$, then we have $\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}]$ for $m \in \{0, 1, 2, 3, 4\}$ and $l \leq u$.*

Proof. Consider the first statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{D}_t = \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\hat{\theta}_n)}_I + \underbrace{\frac{1}{n} \sum_{t=1}^n (\tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n))}_{II}.$$

Focusing on I , we take $\varepsilon > 0$ and let e_1, \dots, e_r denote the unit vectors spanning \mathbb{R}^r . Since $D_t(\theta)$ is continuous in θ we can take $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

$$\mathbb{E}[e'_i D_t] - \varepsilon < \mathbb{E} \left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] \leq \mathbb{E} \left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] < \mathbb{E}[e'_i D_t] + \varepsilon$$

for all $i = 1, \dots, r$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ (Theorem 4.1), we have $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ almost surely. Together with the uniform ergodic theorem we obtain

$$\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) \leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{a.s.} \mathbb{E} \left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] < \mathbb{E}[e'_i D_t] + \varepsilon$$

$$\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) \stackrel{a.s.}{\geq} \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{Y}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{a.s.} \mathbb{E} \left[\inf_{\theta \in \mathcal{Y}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] > \mathbb{E}[e'_i D_t] - \varepsilon.$$

Taking $\varepsilon \searrow 0$ establishes $\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) \xrightarrow{a.s.} \mathbb{E}[e'_i D_t]$ for all i yielding $I \xrightarrow{a.s.} \mathbb{E}[D_t] = \Omega$. Regarding II , we note that for each $\theta \in \Theta$, Assumption 4.4 implies

$$\begin{aligned} & \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\| = \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\ &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \left(\frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right) + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\ &\leq \frac{1}{\tilde{\sigma}_t(\theta)} \left\| \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\tilde{\sigma}_t(\theta)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\ &\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1 \rho^t}{\underline{\omega}} \|D_t(\theta)\| = \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta)\|). \end{aligned} \tag{4.27}$$

We obtain

$$\begin{aligned} \|II\| &\leq \frac{1}{n} \sum_{t=1}^n \|\tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n)\| \leq \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + \|D_t(\hat{\theta}_n)\|) \\ &\stackrel{a.s.}{\leq} \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t). \end{aligned}$$

For each $\varepsilon > 0$, Markov's inequality entails

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^t(1 + U_t) > \varepsilon] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[U_t]}{\varepsilon} = \frac{1 + \mathbb{E}[U_t]}{\varepsilon(1 - \rho)} < \infty$$

since $\rho \in (0, 1)$ and $\mathbb{E}[U_t] < \infty$ by Assumption 4.9(ii). The Borel-Cantelli lemma implies

$$0 = \mathbb{P} \left[\lim_{t \rightarrow \infty} \bigcup_{s=t}^{\infty} \left\{ \rho^s(1 + U_s) > \varepsilon \right\} \right] \geq \mathbb{P} \left[\lim_{t \rightarrow \infty} \rho^t(1 + U_t) > \varepsilon \right] \tag{4.28}$$

and hence $\rho^t(1 + U_t) \rightarrow 0$ almost surely. Cesàro's lemma yields $\frac{1}{n} \sum_{t=1}^n \rho^t(1 + U_t) \xrightarrow{a.s.} 0$ and hence $\|II\| \xrightarrow{a.s.} 0$, which validates the first statement.

Consider the second statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}'_t = \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n)}_{III} + \underbrace{\frac{1}{n} \sum_{t=1}^n \left(\tilde{D}_t(\hat{\theta}_n) \tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) \right)}_{IV}.$$

We focus on *III* and let $\varepsilon > 0$. Since $D_t(\theta)D_t(\theta)'$ is continuous in θ we can take $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

$$\begin{aligned} \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon &< \mathbb{E}\left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j\right] \\ &\leq \mathbb{E}\left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j\right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

for all $i, j = 1, \dots, r$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ by Theorem 4.1, we have $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ almost surely. Together with the uniform ergodic theorem we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \\ &\stackrel{a.s.}{\rightarrow} \mathbb{E}\left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j\right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \\ \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j &\stackrel{a.s.}{\geq} \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \\ &\stackrel{a.s.}{\rightarrow} \mathbb{E}\left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j\right] > \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon \end{aligned}$$

Taking $\varepsilon \searrow 0$ establishes $\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j \xrightarrow{a.s.} \mathbb{E}[e'_i D_t D'_t e_j]$ for all pairs (i, j) yielding *III* $\xrightarrow{a.s.} \mathbb{E}[D_t D'_t] = J$. Consider *IV*; using (4.27) and the elementary inequality

$$\|xx' - yy'\| \leq \|x - y\|^2 + 2\|x - y\| \|y\| \quad (4.29)$$

for all $x, y \in \mathbb{R}^m$, we obtain for $\theta \in \Theta$

$$\begin{aligned} &\left\| \tilde{D}_t(\theta) \tilde{D}'_t(\theta) - D_t(\theta) D'_t(\theta) \right\| \\ &\leq \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\|^2 + 2 \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\| \left\| D_t(\theta) \right\| \\ &\leq \frac{C_1^2}{\underline{\omega}^2} \rho^{2t} \left(1 + \left\| D_t(\theta) \right\| \right)^2 + \frac{2C_1}{\underline{\omega}} \rho^t \left(1 + \left\| D_t(\theta) \right\| \right) \left\| D_t(\theta) \right\| \\ &\leq \frac{C_1^2}{\underline{\omega}^2} \rho^t \left(1 + \left\| D_t(\theta) \right\| \right)^2 + \frac{2C_1}{\underline{\omega}} \rho^t \left(1 + \left\| D_t(\theta) \right\| \right)^2 \\ &= \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left(1 + \left\| D_t(\theta) \right\| \right)^2. \end{aligned} \quad (4.30)$$

Hence, we get

$$\begin{aligned}
 \|IV\| &\leq \frac{1}{n} \sum_{t=1}^n \left\| \tilde{D}_t(\hat{\theta}_n) \tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) \right\| \\
 &\leq \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left(1 + \|D_t(\hat{\theta}_n)\| \right)^2 \\
 &\stackrel{a.s.}{\leq} \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t)^2.
 \end{aligned} \tag{4.31}$$

For each $\varepsilon > 0$, Markov's inequality yields

$$\sum_{t=1}^{\infty} \mathbb{P} \left[\rho^t (1 + U_t)^2 > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^{t/2} \frac{1 + \mathbb{E}[U_t]}{\sqrt{\varepsilon}} = \frac{1 + \mathbb{E}[U_t]}{\sqrt{\varepsilon}(1 - \sqrt{\rho})} < \infty$$

and $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t)^2 \xrightarrow{a.s.} 0$ follows from combining the Borel-Cantelli lemma with Cesàro's lemma. Hence, $\|IV\| \xrightarrow{a.s.} 0$, which validates the second statement.

Consider the third statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{H}_t = \underbrace{\frac{1}{n} \sum_{t=1}^n H_t(\hat{\theta}_n)}_V + \underbrace{\frac{1}{n} \sum_{t=1}^n \left(\tilde{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n) \right)}_{VI}$$

We focus on V and let $\varepsilon > 0$. Since $H_t(\theta)$ is continuous in θ we can take $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

$$\begin{aligned}
 \mathbb{E}[e'_i H_t e_j] - \varepsilon &< \mathbb{E} \left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \right] \\
 &\leq \mathbb{E} \left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \right] < \mathbb{E}[e'_i H_t e_j] + \varepsilon
 \end{aligned}$$

for all $i, j \in \{1, \dots, r\}$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ by Theorem 4.1, we have $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ almost surely. Together with the uniform ergodic theorem we obtain

$$\begin{aligned}
 \frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \xrightarrow{a.s.} \mathbb{E} \left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \right] \\
 &< \mathbb{E}[e'_i H_t e_j] + \varepsilon \\
 \frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j &\stackrel{a.s.}{\geq} \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \xrightarrow{a.s.} \mathbb{E} \left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \right]
 \end{aligned}$$

$$> \mathbb{E}[e'_i H_t e_j] - \varepsilon$$

Taking $\varepsilon \searrow 0$ establishes $\frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j \xrightarrow{a.s.} \mathbb{E}[e'_i H_t e_j]$ for all pairs (i, j) yielding $V \xrightarrow{a.s.} \mathbb{E}[H_t]$. Regarding VI , we note that

$$\begin{aligned} \|\tilde{H}_t(\theta) - H_t(\theta)\| &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \left(\frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right) + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &\leq \frac{1}{\tilde{\sigma}_t(\theta)} \left\| \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\tilde{\sigma}_t(\theta)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1 \rho^t}{\underline{\omega}} \|H_t(\theta)\| = \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|H_t(\theta)\|) \end{aligned} \quad (4.32)$$

for each $\theta \in \Theta$. We obtain

$$\begin{aligned} \|VI\| &\leq \frac{1}{n} \sum_{t=1}^n \|\tilde{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n)\| \leq \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + \|H_t(\hat{\theta}_n)\|) \\ &\stackrel{a.s.}{\leq} \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + V_t). \end{aligned}$$

For each $\varepsilon > 0$, Markov's inequality yields

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^t (1 + V_t) > \varepsilon] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[V_t]}{\varepsilon} = \frac{1 + \mathbb{E}[V_t]}{\varepsilon(1 - \rho)} < \infty$$

and $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + V_t) \xrightarrow{a.s.} 0$ follows from combining the Borel-Cantelli lemma with Cesáro's lemma. Hence, $\|VI\| \xrightarrow{a.s.} 0$, which validates the third statement.

Consider the fourth statement; let $m \in \{0, 1, 2, 3, 4\}$ and take $l, u \in \mathbb{R}$ such that $l \leq u$. We employ the partial integration formula

$$G(b-)H(b-) - G(a-)H(a-) = \int_{[a,b)} G(t-)dH(t) + \int_{[a,b)} H(s)dG(s) \quad (4.33)$$

with G and H both right-continuous functions being locally of bounded variation and $a \leq t < b$ to expand

$$\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} - \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}] = \int_{[l,u)} x^m d\hat{\mathbb{F}}_n(x) - \int_{[l,u)} x^m dF(x)$$

$$= u^m \left(\hat{\mathbb{F}}_n(u-) - F(u) \right) - l^m \left(\hat{\mathbb{F}}_n(l-) - F(l) \right) + \int_{[l,u)} \left(\hat{\mathbb{F}}_n(x) - F(x) \right) dx^m.$$

Lemma 4.1 implies $\hat{\mathbb{F}}_n(u-) \xrightarrow{a.s.} F(u)$ and $\hat{\mathbb{F}}_n(l-) \xrightarrow{a.s.} F(l)$ and together with the dominated convergence theorem yields $\int_{[l,u)} (\hat{\mathbb{F}}_n(x) - F(x)) dx^m \xrightarrow{a.s.} 0$. Thus,

$$\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}]$$

for $m \in \{0, 1, 2, 3, 4\}$ and $l, u \in \mathbb{R}$. Since $\mathbb{E}[|\eta_t|^m] < \infty$ and $\mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}] = \int_l^u x^m f(x) dx$ is continuous in l and u it is easy to see that the result extends to $l = -\infty$ and $u = \infty$, which validates the fourth statement and completes the proof. \square

Lemma 4.3. *Suppose Assumptions 4.1–4.9 hold with $a = \pm 6$, $b = 6$ and $c = 2$ and let $\mathcal{I}_n = (\xi_\alpha - a_n, \xi_\alpha + a_n)$ with $a_n \sim n^{-\varrho} \log n$ for some $\varrho \in (0, 1)$. Then, we have*

$$\sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n} (\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n} (F(x) - F(y)) \right| \xrightarrow{P} 0.$$

Replacing any $\hat{\mathbb{F}}_n(\cdot)$ by $\hat{\mathbb{F}}_n(\cdot -)$ does not alter the result.

Proof. We follow Berkes and Horváth (2003) and define

$$\begin{aligned} \tilde{\gamma}_t(u) &= \tilde{\sigma}_t(\theta_0 + n^{-1/2}u) / \sigma_t(\theta_0) \\ \gamma_t(u) &= \sigma_t(\theta_0 + n^{-1/2}u) / \sigma_t(\theta_0) \\ \zeta_t(x, u) &= \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - F(x \tilde{\gamma}_t(u)) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \\ S_n(x, u) &= \sum_{t=1}^n \zeta_t(x, u) \\ \mathbb{F}_n(x) &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}} \end{aligned}$$

for $x \in \mathbb{R}$ and $u \in \mathbb{R}^r$, where we suppress the dependence of $\tilde{\gamma}_t(u)$, $\gamma_t(u)$ and $\zeta_t(x, u)$ on n for notational simplicity. Let $A > 0$ and write $\mathcal{V}(\xi_\alpha)$ to denote the neighborhood around ξ_α on which f is continuous; see Assumption 4.5(ii). Since $\xi_\alpha < 0$, we can take a compact neighborhood $\mathcal{X} = [\underline{x}, \bar{x}] \subset \mathcal{V}(\xi_\alpha)$ such that $\xi_\alpha \in \mathcal{X}$ and $\bar{x} < 0$. We establish the result in seven steps:

Step 1: $\mathbb{E}[|S_n(x, u)|^4] = O(n)$ for all $x \in \mathcal{X}$ and for all $u \in \{u \in \mathbb{R}^r : \|u\| \leq A\}$;

Step 2: $\sup_{x \in \mathcal{X}} |S_n(x, u)| = o_p(\sqrt{n})$ for all $u \in \{u \in \mathbb{R}^r : \|u\| \leq A\}$;

Step 3: $\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| = o_p(\sqrt{n})$;

Step 4: $\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n (F(x\tilde{\gamma}_t(u)) - F(x)) - xf(x)\Omega'u \right| = o_p(1)$;

Step 5: $\sup_{x \in \mathcal{X}} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \mathbb{F}_n(x)) - xf(x)\Omega'\sqrt{n}(\hat{\theta}_n - \theta_0) \right| = o_p(1)$;

Step 6: $\sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| = O(n^{-\ell/2} \log n)$ a.s.;

Step 7: $\sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \xrightarrow{P} 0$.

Step 1 to *Step 5* are similar to the proofs of Berkes and Horváth (2003), whereas *Step 6* resembles Bahadur (1966, Lemma 1). Throughout *Step 1* to *Step 4* we take $\delta \in (0, 1/2)$ such that $\mathcal{X}_\delta = [\underline{x}(1 + 2\delta), \bar{x}(1 - 2\delta)]$ satisfies $\mathcal{X} \subset \mathcal{X}_\delta \subset \mathcal{V}(\xi_\alpha)$. Because f is continuous on \mathcal{X}_δ and \mathcal{X}_δ is compact, f is uniformly continuous on \mathcal{X}_δ and there exists a finite $M > 0$ such that

$$\sup_{x \in \mathcal{X}_\delta} f(x) \leq M. \quad (4.34)$$

Consider *Step 1*; let \mathcal{F}_t be the σ -algebra generated by $\zeta_t, \zeta_{t-1}, \dots$ and note that $\{S_t(x, u), \mathcal{F}_t\}$ is a martingale given x and u . Theorem 2.11 of Hall and Heyde (1980) yields

$$\mathbb{E}[|S_n(x, u)|^4] \leq C \left(\mathbb{E} \left[\max_{1 \leq t \leq n} \zeta_t^4(x, u) \right] + \mathbb{E} \left[\left(\sum_{t=1}^n \mathbb{E}_{t-1} [\zeta_t^2(x, u)] \right)^2 \right] \right),$$

for some absolute constant $C > 0$ independent of x and u , where $\mathbb{E}_{t-1} = \mathbb{E}[\cdot | \mathcal{F}_{t-1}]$ is the expectation given \mathcal{F}_{t-1} . Since $|\zeta_t(x, u)| \leq 2$ for all $t = 1, \dots, n$ such that $\mathbb{E}[\max_{1 \leq t \leq n} \zeta_t^4(x, u)] \leq 16$, it suffices to show that

$$\mathbb{E} \left[\left(\sum_{t=1}^n \mathbb{E}_{t-1} [\zeta_t^2(x, u)] \right)^2 \right] = O(n). \quad (4.35)$$

First, we focus on the inner part $\mathbb{E}_{t-1}[\zeta_t^2(x, u)]$ and decompose $\zeta_t(x, u)$ into

$$\zeta_t(x, u) = \zeta_{t,1}(x, u) + \zeta_{t,2}(x, u)$$

with

$$\begin{aligned}\zeta_{t,1}(x, u) &= \mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} - F(x\tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}} + F(x\gamma_t(u)) \\ \zeta_{t,2}(x, u) &= \mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}} - F(x\gamma_t(u)) - \mathbb{1}_{\{\eta_t \leq x\}} + F(x).\end{aligned}$$

The elementary inequality

$$\left(\sum_{i=1}^m x_i\right)^2 \leq m \sum_{i=1}^m x_i^2 \quad (4.36)$$

for all $x_1, \dots, x_m \in \mathbb{R}$ with $m \in \mathbb{N}$ implies that

$$\mathbb{E}_{t-1}[\zeta_t^2(x, u)] \leq 2\left(\mathbb{E}_{t-1}[\zeta_{t,1}^2(x, u)] + \mathbb{E}_{t-1}[\zeta_{t,2}^2(x, u)]\right).$$

Moreover, the inequality $\text{Var}[\mathbb{1}_{\{X \leq y\}} - \mathbb{1}_{\{X \leq z\}}] \leq |F_X(y) - F_X(z)|$ for $y, z \in \mathbb{R}$ and $X \sim F_X$ gives

$$\begin{aligned}\mathbb{E}_{t-1}[\zeta_{t,1}^2(x, u)] &= \text{Var}_{t-1}[\mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}}] \\ &\leq |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{t-1}[\zeta_{t,2}^2(x, u)] &= \text{Var}_{t-1}[\mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\}}] \\ &\leq |F(x\gamma_t(u)) - F(x)|.\end{aligned}$$

Combining results, it follows that

$$\mathbb{E}_{t-1}[\zeta_t^2(x, u)] \leq 2\left(|F(x\gamma_t(u)) - F(x)| + |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\right). \quad (4.37)$$

Employing (4.37), we obtain that the left-hand side in (4.35) is bounded by

$$4\mathbb{E}\left[\left(\sum_{t=1}^n |F(x\gamma_t(u)) - F(x)| + \sum_{t=1}^n |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\right)^2\right]$$

$$\leq 8 \underbrace{\left(\mathbb{E} \left[\left(\sum_{t=1}^n |F(x\gamma_t(u)) - F(x)| \right)^2 \right] \right)}_I + \underbrace{\mathbb{E} \left[\left(\sum_{t=1}^n |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))| \right)^2 \right]}_{II},$$

where the last inequality follows from applying (4.36) once more. It suffices to show that both terms are $O(n)$. Consider I ; The Cauchy-Schwarz inequality yields

$$\begin{aligned} I &= \sum_{t=1}^n \sum_{\tau=1}^n \mathbb{E} \left[|F(x\gamma_t(u)) - F(x)| |F(x\gamma_\tau(u)) - F(x)| \right] \\ &\leq \sum_{t=1}^n \sum_{\tau=1}^n \left(\mathbb{E} \left[(F(x\gamma_t(u)) - F(x))^2 \right] \right)^{\frac{1}{2}} \left(\mathbb{E} \left[(F(x\gamma_\tau(u)) - F(x))^2 \right] \right)^{\frac{1}{2}}. \end{aligned} \quad (4.38)$$

Henceforth, we take n sufficiently large such that $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$. The mean value theorem implies

$$\begin{aligned} \sup_{\|u\| \leq A} |\gamma_t(u) - 1| &= \sup_{\|u\| \leq A} \left| \frac{\sigma_t(\theta_0 + u/\sqrt{n}) - \sigma_t(\theta_0)}{\sigma_t(\theta_0)} \right| \\ &= \sup_{\|u\| \leq A} \left| \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta'} \frac{1}{\sqrt{n}} u \right| = \frac{1}{\sqrt{n}} \sup_{\|u\| \leq A} \left| \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} D'_t(\bar{\theta}_n) u \right| \\ &\leq \frac{1}{\sqrt{n}} \sup_{\sqrt{n}\|\theta - \theta_0\| \leq A} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\sqrt{n}\|\theta - \theta_0\| \leq A} \|D_t(\theta)\| \sup_{\|u\| \leq A} \|u\| \leq \frac{A}{\sqrt{n}} T_t U_t, \end{aligned} \quad (4.39)$$

where T_t and U_t are defined in (4.25) and $\bar{\theta}_n$ lies between θ_0 and $\theta_0 + u/\sqrt{n}$. Define the event

$$\mathcal{A}_{n,t} = \left\{ \frac{A}{\sqrt{n}} T_t U_t \leq \delta \right\}, \quad (4.40)$$

where δ is given in the text preceding (4.34). The inner term of (4.38) can be bounded by

$$\begin{aligned} &\mathbb{E} \left[(F(x\gamma_t(u)) - F(x))^2 \right] \\ &= \mathbb{E} \left[\underbrace{(F(x\gamma_t(u)) - F(x))^2}_{\leq 1} (\mathbb{1}_{\{\mathcal{A}_{n,t}^c\}} + \mathbb{1}_{\{\mathcal{A}_{n,t}\}}) \right] \\ &\leq \underbrace{\mathbb{P}[\mathcal{A}_{n,t}^c]}_{I_1} + \underbrace{\mathbb{E} \left[(F(x\gamma_t(u)) - F(x))^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}} \right]}_{I_2}, \end{aligned} \quad (4.41)$$

where the superscript c denotes the event's complement. Using Markov's inequality, the Cauchy-Schwarz inequality and Assumption 4.9, I_1 can be bounded by

$$I_1 = \mathbb{P}\left[\frac{A}{\sqrt{n}}T_t U_t > \delta\right] \leq \frac{A^2}{n\delta^2}\mathbb{E}[T_t^2 U_t^2] \leq \frac{A^2}{n\delta^2}\underbrace{\left(\mathbb{E}[T_t^4]\right)}_{<\infty}^{\frac{1}{2}}\underbrace{\left(\mathbb{E}[U_t^4]\right)}_{<\infty}^{\frac{1}{2}} \quad (4.42)$$

and, thus, $I_1 = O(n^{-1})$. Regarding I_2 , the mean value theorem implies

$$I_2 = \mathbb{E}\left[x^2 f^2(x\bar{\gamma}_t)(\gamma_t(u) - 1)^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}}\right]$$

with $\bar{\gamma}_t$ being between $\gamma_t(u)$ and 1. Since $|\bar{\gamma}_t - 1| \leq |\gamma_t(u) - 1| \leq \delta$ in the event of $\mathcal{A}_{n,t}$, we have $x\bar{\gamma}_t \in \mathcal{X}_\delta$. Employing (4.34), (4.39), the Cauchy-Schwarz inequality and Assumption 4.9, we establish

$$I_2 \leq \mathbb{E}\left[\underline{x}^2 M^2 \frac{A^2}{n} T_t^2 U_t^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}}\right] \leq \frac{\underline{x}^2 M^2 A^2}{n} \underbrace{\left(\mathbb{E}[T_t^4]\right)}_{<\infty}^{\frac{1}{2}} \underbrace{\left(\mathbb{E}[U_t^4]\right)}_{<\infty}^{\frac{1}{2}} \quad (4.43)$$

and $I_2 = O(n^{-1})$ follows. Combining (4.41) to (4.43) yields

$$\mathbb{E}\left[\left(F(x\gamma_t(u)) - F(x)\right)^2\right] \leq I_1 + I_2 = O(n^{-1})$$

and, together with (4.38), we get

$$I \leq \sum_{t=1}^n \sum_{r=1}^n O(n^{-1/2}) O(n^{-1/2}) = O(n).$$

Next, we consider II , which can be bounded analogously to (4.38) by

$$\begin{aligned} II &\leq \sum_{t=1}^n \sum_{\tau=1}^n \left(\mathbb{E}\left[\left(F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))\right)^2\right] \right)^{\frac{1}{2}} \\ &\quad \times \left(\mathbb{E}\left[\left(F(x\tilde{\gamma}_\tau(u)) - F(x\gamma_\tau(u))\right)^2\right] \right)^{\frac{1}{2}}. \end{aligned} \quad (4.44)$$

Assumption 4.4(i) gives

$$\begin{aligned} & \sup_{||u|| \leq A} |\tilde{\gamma}_t(u) - \gamma_t(u)| \\ &= \sup_{||u|| \leq A} \frac{|\tilde{\sigma}_t(\theta_0 + n^{-1/2}u) - \sigma_t(\theta_0 + n^{-1/2}u)|}{\sigma_t(\theta_0)} \leq \rho^t \frac{C_1}{\underline{\omega}}. \end{aligned} \quad (4.45)$$

We define the events

$$\mathcal{B}_t = \left\{ \rho^t \frac{C_1}{\underline{\omega}} \leq \delta \rho^{t/2} \right\} \quad \text{and} \quad \mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t. \quad (4.46)$$

In analogy to (4.41), the inner part of (4.44) can be bounded by

$$\begin{aligned} & \mathbb{E} \left[\left(F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \\ & \leq \underbrace{\mathbb{P}[\mathcal{C}_{n,t}^c]}_{II_1} + \underbrace{\mathbb{E} \left[\left(F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \right]}_{II_2}. \end{aligned}$$

Employing (4.42) and Markov's inequality yields

$$\begin{aligned} II_1 &= \mathbb{P}[\mathcal{A}_{n,t}^c \cup \mathcal{B}_t^c] \leq \mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P}[\mathcal{B}_t^c] = \mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P} \left[\rho^{t/2} \frac{C_1}{\underline{\omega}} > \delta \right] \\ &\leq \frac{A^2}{n\delta^2} \left(\mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left(\mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + (\rho^{s/2})^t \frac{\mathbb{E}[C_1^s]}{\delta^s \underline{\omega}^s} = O(n^{-1}) + O((\rho^{s/2})^t). \end{aligned} \quad (4.47)$$

Regarding II_2 , the mean value theorem implies

$$II_2 = \mathbb{E} \left[x^2 f^2(x\check{\gamma}_t) (\tilde{\gamma}_t(u) - \gamma_t(u))^2 \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \right]$$

with $\check{\gamma}_t$ between $\tilde{\gamma}_t(u)$ and $\gamma_t(u)$. Since

$$|\check{\gamma}_t - 1| \leq |\check{\gamma}_t - \gamma_t(u)| + |\gamma_t(u) - 1| \leq |\tilde{\gamma}_t(u) - \gamma_t(u)| + |\gamma_t(u) - 1| \leq 2\delta$$

in the event of $\mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t$, we have $x\check{\gamma}_t \in \mathcal{X}_\delta$. Employing (4.34) and (4.45) we obtain

$$II_2 \leq \mathbb{E} \left[\underbrace{x^2 M^2 \left(\rho^t \frac{C_1}{\underline{\omega}} \right)^2}_{\leq \delta^2 \rho^t} \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \right] \leq x^2 M^2 \delta^2 \rho^t = O(\rho^t). \quad (4.48)$$

Equations (4.47) and (4.48) imply

$$\mathbb{E} \left[\left(F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \leq C(n^{-1} + \rho^t + (\rho^{s/2})^t)$$

for some constant $C > 0$. Inserting this result into (4.44), we conclude

$$II \leq C \sum_{t=1}^n \sum_{\tau=1}^n \left(n^{-1} + \rho^t + (\rho^{s/2})^t \right)^{\frac{1}{2}} \left(n^{-1} + \rho^\tau + (\rho^{s/2})^\tau \right)^{\frac{1}{2}} = O(n),$$

which completes *Step 1*.

In *Step 2* we divide \mathcal{X} into intervals with the points $\underline{x} = x_1 < x_2 < \dots < x_N < x_{N+1} = \bar{x}$ satisfying $0.5 n^{-3/4} \leq x_{j+1} - x_j \leq n^{-3/4}$ for all $j = 1, \dots, N$ and $N \in \mathbb{N}$. It follows that $N = O(n^{3/4})$. We obtain

$$\begin{aligned} \sup_{x \in \mathcal{X}} |S_n(x, u)| &= \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u)| \\ &\leq \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} \left(|S_n(x_{j+1}, u)| + |S_n(x, u) - S_n(x_{j+1}, u)| \right) \\ &\leq \max_{1 \leq j \leq N} |S_n(x_{j+1}, u)| + \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)|. \end{aligned} \quad (4.49)$$

We bound the second term using the elementary inequality

$$|x - y| \leq \max\{x, y\} \quad (4.50)$$

for all $x, y \geq 0$. For $j = 1, \dots, N$, we have

$$\begin{aligned} &\sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)| \\ &= \sup_{x_j \leq x \leq x_{j+1}} \left| \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x\}} + F(x_{j+1}\tilde{\gamma}_t(u)) - F(x\tilde{\gamma}_t(u)) \right) \right. \\ &\quad \left. - \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x) \right) \right| \\ &\leq \sup_{x_j \leq x \leq x_{j+1}} \max \left\{ \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x\}} + F(x_{j+1}\tilde{\gamma}_t(u)) - F(x\tilde{\gamma}_t(u)) \right), \right. \\ &\quad \left. \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x) \right) \right\} \end{aligned}$$

$$\leq \max \left\{ \underbrace{\sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_{j+1}\tilde{\gamma}_t(u)) - F(x_j\tilde{\gamma}_t(u)) \right)}_{=A_n}, \quad (4.51) \right. \\ \left. \underbrace{\sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x_j\tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x_j) \right)}_{=B_n} \right\}.$$

Note that A_n and B_n are positive, where the later can be rewritten as

$$\begin{aligned} B_n &= \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\tilde{\gamma}_t(u)\}} - F(x_{j+1}\tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} + F(x_{j+1}) \right) \\ &\quad - \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_j\tilde{\gamma}_t(u)\}} - F(x_j\tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_j) \right) \quad (4.52) \\ &\quad + \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_{j+1}\tilde{\gamma}_t(u)) - F(x_j\tilde{\gamma}_t(u)) \right) \\ &= S_n(x_{j+1}, u) - S_n(x_j, u) + A_n. \end{aligned}$$

It follows from (4.51) and (4.52) that

$$\sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)| \leq |S_n(x_{j+1}, u)| + |S_n(x_j, u)| + A_n. \quad (4.53)$$

Moreover, A_n expands as follows:

$$\begin{aligned} A_n &= \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - F(x_{j+1}) - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_j) \right) + n(F(x_{j+1}) - F(x_j)) \\ &\quad + \sum_{t=1}^n \left(F(x_{j+1}\tilde{\gamma}_t(u)) - F(x_j\tilde{\gamma}_t(u)) \right) \quad (4.54) \end{aligned}$$

Using equations (4.49), (4.53) and (4.54), we establish

$$\sup_{x \in \mathcal{X}} |S_n(x, u)| \leq 3III + IV + V + VI + 2VII, \quad (4.55)$$

where

$$\begin{aligned} III &= \max_{1 \leq j \leq N+1} |S_n(x_j, u)| \\ IV &= \max_{1 \leq j \leq N} n(F(x_{j+1}) - F(x_j)) \end{aligned}$$

$$\begin{aligned}
 V &= \max_{1 \leq j \leq N} \left| \sum_{t=1}^n (\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - F(x_{j+1})) - \sum_{t=1}^n (\mathbb{1}_{\{\eta_t \leq x_j\}} - F(x_j)) \right| \\
 VI &= \max_{1 \leq j \leq N} \sum_{t=1}^n \left(F(x_{j+1} \gamma_t(u)) - F(x_j \gamma_t(u)) \right) \\
 VII &= \max_{1 \leq j \leq N+1} \sum_{t=1}^n \left| F(x_j \tilde{\gamma}_t(u)) - F(x_j \gamma_t(u)) \right|.
 \end{aligned}$$

We look at each term in turn. For each $\varepsilon > 0$, Markov's inequality implies

$$\begin{aligned}
 \mathbb{P}[III \geq \sqrt{n}\varepsilon] &= \mathbb{P}\left[\max_{1 \leq j \leq N+1} |S_n(x_j, u)|^4 \geq n^2 \varepsilon^4\right] \\
 &\leq \frac{1}{n^2 \varepsilon^4} \mathbb{E}\left[\max_{1 \leq j \leq N+1} |S_n(x_j, u)|^4\right] \leq \sum_{j=1}^{N+1} \frac{1}{n^2 \varepsilon^4} \mathbb{E}[|S_n(x_j, u)|^4] \rightarrow 0
 \end{aligned}$$

as $N = O(n^{3/4})$ and $\mathbb{E}[|S_n(x, u)|^4] = O(n)$ by *Step 1*. Thus, we have $III = o_p(\sqrt{n})$. Regarding *IV*, the mean value theorem and (4.34) yield

$$F(x_{j+1}) - F(x_j) = f(\check{x}_j)(x_{j+1} - x_j) \leq Mn^{-3/4}, \quad (4.56)$$

where $\check{x}_j \in (x_j, x_{j+1})$. It follows that

$$IV \leq nMn^{-3/4} = Mn^{1/4}$$

yielding $IV = O(n^{1/4})$. Further, Theorem 4.3.1 of Csörgő and Révész (1981) implies that there exists a sequence of Brownian bridges $\{B_n(y) : 0 \leq y \leq 1\}$ such that

$$\begin{aligned}
 V/\sqrt{n} &= \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_{j+1}) - F(x_{j+1})) - \sqrt{n}(\mathbb{F}_n(x_j) - F(x_j)) \right| \\
 &\leq \max_{1 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| \\
 &\quad + \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_j) - F(x_j)) - B_n(F(x_j)) \right| \\
 &\quad + \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_{j+1}) - F(x_{j+1})) - B_n(F(x_{j+1})) \right| \\
 &\leq \max_{1 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| \\
 &\quad + 2 \sup_{x \in \mathbb{R}} \left| \sqrt{n}(\mathbb{F}_n(x) - F(x)) - B_n(F(x)) \right|
 \end{aligned}$$

$$\stackrel{a.s.}{=} \max_{1 \leq j \leq N} \left| \underbrace{B_n(F(x_{j+1})) - B_n(F(x_j))}_{Z_{n,j}} \right| + o(1).$$

Next, we show that $\max_{0 \leq j \leq N} |Z_{n,j}| = o_p(1)$. By the definition of a Brownian bridge (cf. Csörgő and Révész, 1981, p. 41), $Z_{n,j}$ is Gaussian with mean 0 and variance

$$\mathbb{V}\text{ar}[Z_{n,j}] = (F(x_{j+1}) - F(x_j)) \underbrace{\left(1 - (F(x_{j+1}) - F(x_j))\right)}_{\leq 1} \leq Mn^{-3/4}$$

by (4.56). In addition, we have $\mathbb{E}[Z_{n,j}^4] = 3(\mathbb{V}\text{ar}[Z_{n,j}])^2 \leq 3M^2n^{-3/2}$. Thus, for each $\varepsilon > 0$, Markov's inequality implies

$$\begin{aligned} \mathbb{P}\left[\max_{1 \leq j \leq N} |Z_{n,j}| \geq \varepsilon\right] &= \mathbb{P}\left[\max_{1 \leq j \leq N} Z_{n,j}^4 \geq \varepsilon^4\right] \leq \frac{1}{\varepsilon^4} \mathbb{E}\left[\max_{1 \leq j \leq N} Z_{n,j}^4\right] \\ &\leq \frac{1}{\varepsilon^4} \mathbb{E}\left[\sum_{j=1}^N Z_{n,j}^4\right] \leq \frac{1}{\varepsilon^4} \sum_{j=1}^N 3M^2n^{-3/2} = \frac{3M^2}{\varepsilon^4} n^{-3/2} N \rightarrow 0 \end{aligned}$$

as $N = O(n^{3/4})$ and we conclude $\max_{1 \leq j \leq N} |Z_{n,j}| = o_p(1)$. Thus, $V = o_p(\sqrt{n})$. In analogy to (4.41), we bound VI by

$$VI \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}}_{VI_1} + \underbrace{\max_{1 \leq j \leq N} \sum_{t=1}^n \left(F(x_{j+1}\gamma_t(u)) - F(x_j\gamma_t(u))\right) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}}_{VI_2}. \quad (4.57)$$

Concerning the first subterm, for each $\varepsilon > 0$, Markov's inequality and (4.42) yield

$$\begin{aligned} \mathbb{P}[VI_1 \geq \sqrt{n}\varepsilon] &\leq \frac{1}{\sqrt{n}\varepsilon} \mathbb{E}\left[\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}\right] = \frac{1}{\sqrt{n}\varepsilon} \sum_{t=1}^n \mathbb{P}[\mathcal{A}_{n,t}^c] \\ &\leq \frac{A^2}{\sqrt{n}\varepsilon\delta^2} \left(\mathbb{E}[T_t^4]\right)^{\frac{1}{2}} \left(\mathbb{E}[U_t^4]\right)^{\frac{1}{2}} = O(n^{-1/2}). \end{aligned} \quad (4.58)$$

Thus, we have $VI_1 = o_p(\sqrt{n})$. Regarding VI_2 , the mean value theorem implies

$$VI_2 = \max_{1 \leq j \leq N} \sum_{t=1}^n \gamma_t(u) f(\tilde{x}_j \gamma_t(u)) (x_{j+1} - x_j) \mathbb{1}_{\{\mathcal{A}_{n,t}\}},$$

where \tilde{x}_j lies between x_j and x_{j+1} . Since $|\gamma_t(u) - 1| \leq \delta$ in the event of $\mathcal{A}_{n,t}$, we

have $\tilde{x}_j \gamma_t(u) \in \mathcal{X}_\delta$. Employing (4.34) and (4.39), we get

$$VI_2 \leq \sum_{t=1}^n \left(1 + \frac{A}{\sqrt{n}} T_t U_t \right) M n^{-3/4} = M n^{1/4} + \frac{A}{n^{1/4}} \frac{1}{n} \sum_{t=1}^n T_t U_t$$

While the first term is of order $O(n^{1/4})$, the second vanishes almost surely as

$$\frac{1}{n} \sum_{t=1}^n T_t U_t \leq \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[T_t^2] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[U_t^2] < \infty} \quad (4.59)$$

by Markov's inequality, the uniform ergodic theorem and Assumption 4.9. Hence, $VI_2 = O(n^{1/4})$ almost surely. Next, we show

$$VII^\diamond = \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x \tilde{\gamma}_t(u)) - F(x \gamma_t(u)) \right| = O_p(1), \quad (4.60)$$

which implies $VII = O_p(1)$. Similar to (4.41), we bound VII^\diamond by

$$VII^\diamond \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{C}_{n,t}^c\}}}_{VII_1^\diamond} + \underbrace{\sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x_j \tilde{\gamma}_t(u)) - F(x_j \gamma_t(u)) \right| \mathbb{1}_{\{\mathcal{C}_{n,t}\}}}_{VII_2^\diamond}$$

where the event $\mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t$ is defined in (4.46). We show that both terms are $O_p(1)$. Employing Markov's inequality and (4.47), we have for each $C > 0$

$$\begin{aligned} \mathbb{P}[VII_1^\diamond \geq C] &\leq \frac{1}{C} \mathbb{E}[VII_1^\diamond] = \frac{1}{C} \sum_{t=1}^n \mathbb{P}[\mathcal{C}_{n,t}^c] \leq \frac{1}{C} \sum_{t=1}^n \left(\mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P}[\mathcal{B}_t^c] \right) \\ &\leq \frac{1}{C} \sum_{t=1}^n \left(\frac{A^2}{n \delta^2} \left(\mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left(\mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + (\rho^{s/2})^t \frac{\mathbb{E}[C_1^s]}{\delta^s \underline{\omega}^s} \right) \\ &\leq \frac{1}{C} \left(\frac{A^2}{\delta^2} \left(\mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left(\mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + \frac{\mathbb{E}[C_1^s]}{\underline{\omega}^s \delta^s (1 - \rho^{s/2})} \right). \end{aligned} \quad (4.61)$$

Choosing C sufficiently large, $\mathbb{P}[VII_1^\diamond \geq C]$ can be made sufficiently small and we

conclude $V_1^\diamond = O_p(1)$. Analogously to (4.48) we obtain

$$\begin{aligned} VII_2^\diamond &= \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| x f(x \check{\gamma}_t) (\tilde{\gamma}_t(u) - \gamma_t(u)) \right| \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \\ &\leq \sum_{t=1}^n |\underline{x}| M \underbrace{\frac{C_1 \rho^t}{\omega} \mathbb{1}_{\{\mathcal{C}_{n,t}\}}}_{\leq \delta \rho^{t/2}} \leq \sum_{t=1}^n |\underline{x}| M \delta \rho^{t/2} \leq \frac{2|\underline{x}| M \delta}{(1 - \sqrt{\rho})^2} = O(1) \end{aligned} \quad (4.62)$$

and we conclude $VII^\diamond = O_p(1)$. *Step 2* is completed.

In *Step 3* we divide the (hyper-)cube $[-A, A]^r$ into $L = (2N)^r$ cubes with side length A/N and $N \in \mathbb{N}$. In case of a cube ℓ , $u_\bullet(\ell)$ and $u^\bullet(\ell)$ denote the lower left and upper right vertex of ℓ .¹¹ Similar to (4.49), we obtain

$$\begin{aligned} \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| &\leq \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u^\bullet(\ell))| \\ &\quad + \max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \sup_{x \in \mathcal{X}} |S_n(x, u) - S_n(x, u^\bullet(\ell))|. \end{aligned} \quad (4.63)$$

We focus on the second term. Fix $\ell \in \{1, \dots, L\}$ and consider u satisfying $u_\bullet(\ell) \leq u \leq u^\bullet(\ell)$ (element-by-element comparison). Assumption 4.8 implies $\tilde{\gamma}_t(u_\bullet(\ell)) \leq \tilde{\gamma}_t(u) \leq \tilde{\gamma}_t(u^\bullet(\ell))$. Since $x < 0$ for all $x \in \mathcal{X}$, the elementary inequality (4.50) implies

$$\begin{aligned} &|S_n(x, u) - S_n(x, u^\bullet(\ell))| \\ &= \left| \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - F(x \tilde{\gamma}_t(u)) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \right. \\ &\quad \left. - \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x \tilde{\gamma}_t(u^\bullet(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \right| \\ &= \left| \underbrace{\sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} \right)}_{\geq 0} - \underbrace{\sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right)}_{\geq 0} \right| \\ &\leq \max \left\{ \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} \right), \right. \\ &\quad \left. \sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right) \right\} \end{aligned} \quad (4.64)$$

¹¹ Lower left (right) vertex means that all coordinates of $u_\bullet(\ell)$ ($u^\bullet(\ell)$) are less (larger) than or equal to the corresponding coordinates of any elements of ℓ .

$$\leq \max \left\{ \underbrace{\sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u_{\bullet}(\ell))\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^{\bullet}(\ell))\}} \right)}_{=C_n}, \right. \\ \left. \underbrace{\sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u_{\bullet}(\ell))) - F(x \tilde{\gamma}_t(u^{\bullet}(\ell))) \right)}_{=D_n} \right\}.$$

Note that C_n can be written as

$$\begin{aligned} C_n &= \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u_{\bullet}(\ell))\}} - F(x \tilde{\gamma}_t(u_{\bullet}(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \\ &\quad - \sum_{t=1}^n \left(\mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^{\bullet}(\ell))\}} - F(x \tilde{\gamma}_t(u^{\bullet}(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \\ &\quad + \sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u_{\bullet}(\ell))) - F(x \tilde{\gamma}_t(u^{\bullet}(\ell))) \right) \\ &= S_n(x, u_{\bullet}(\ell)) - S_n(x, u^{\bullet}(\ell)) + D_n. \end{aligned} \quad (4.65)$$

Combining (4.64) and (4.65), we find

$$|S_n(x, u) - S_n(x, u^{\bullet}(\ell))| \leq |S_n(x, u_{\bullet}(\ell))| + |S_n(x, u^{\bullet}(\ell))| + |D_n|. \quad (4.66)$$

Moreover, D_n expands as follows:

$$\begin{aligned} D_n &= \sum_{t=1}^n \left(F(x \gamma_t(u_{\bullet}(\ell))) - F(x \gamma_t(u^{\bullet}(\ell))) \right) \\ &\quad + \sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u_{\bullet}(\ell))) - F(x \gamma_t(u_{\bullet}(\ell))) \right) \\ &\quad - \sum_{t=1}^n \left(F(x \tilde{\gamma}_t(u^{\bullet}(\ell))) - F(x \gamma_t(u^{\bullet}(\ell))) \right) \end{aligned} \quad (4.67)$$

Equations (4.63) and (4.67) lead to

$$\sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| \leq 2VIII + IX + X + XI + XII \quad (4.68)$$

with

$$VIII = \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u^{\bullet}(\ell))|$$

$$\begin{aligned}
 IX &= \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u_\bullet(\ell))| \\
 X &= \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x\tilde{\gamma}_t(u_\bullet(\ell))) - F(x\gamma_t(u_\bullet(\ell))) \right| \\
 XI &= \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x\tilde{\gamma}_t(u^\bullet(\ell))) - F(x\gamma_t(u^\bullet(\ell))) \right| \\
 XII &= \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left(F(x\gamma_t(u_\bullet(\ell))) - F(x\gamma_t(u^\bullet(\ell))) \right).
 \end{aligned}$$

$VIII$ and IX are $o_p(\sqrt{n})$ for fixed L by *Step 2* whereas $X = O_p(1)$ and $XI = O_p(1)$ by (4.60). In analogy to (4.41), we bound XII by

$$XII \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}}_{XII_1} + \underbrace{\max_{1 \leq j \leq N} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left(F(x\gamma_t(u_\bullet(\ell))) - F(x\gamma_t(u^\bullet(\ell))) \right) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}}_{XII_2}.$$

We have $XII_1 = o_p(\sqrt{n})$ by (4.58). Regarding XII_2 , the mean value theorem implies

$$XII_2 = \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n x f(x\tilde{\gamma}_t)(\gamma_t(u_\bullet(\ell)) - \gamma_t(u^\bullet(\ell))) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}$$

with $\tilde{\gamma}_t$ lying between $\gamma_t(u_\bullet(\ell))$ and $\gamma_t(u^\bullet(\ell))$. Since $|\tilde{\gamma}_t - 1| \leq 2\delta$ in the event of $\mathcal{A}_{n,t}$, we have $x\tilde{\gamma}_t \in \mathcal{X}_\delta$ for all $x \in \mathcal{X}$. Taking n sufficiently large such that $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$, (4.34) and the mean value theorem (see Assumptions 4.3 and 4.4) imply

$$\begin{aligned}
 XII_2 &\leq |\underline{x}| M \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n (\gamma_t(u^\bullet(\ell)) - \gamma_t(u_\bullet(\ell))) \\
 &= |\underline{x}| M \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{\sigma_t(\theta_0 + n^{-1/2}u^\bullet(\ell)) - \sigma_t(\theta_0 + n^{-1/2}u_\bullet(\ell))}{\sigma_t(\theta_0)} \\
 &= |\underline{x}| M \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta'} \frac{1}{\sqrt{n}} (u^\bullet(\ell) - u_\bullet(\ell)) \\
 &\leq \frac{|\underline{x}| M}{\sqrt{n}} \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} \left\| \frac{1}{\sigma_t(\bar{\theta}_n)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta} \right\| \|u^\bullet(\ell) - u_\bullet(\ell)\| \\
 &\leq \frac{rA|\underline{x}|M}{\sqrt{n}N} \sum_{t=1}^n \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \|D_t(\theta)\|
 \end{aligned}$$

$$\leq \frac{rA|\underline{x}|M}{\sqrt{n}N} \sum_{t=1}^n T_t U_t,$$

where $\theta_0 + n^{-1/2}u_{\bullet}(\ell) \leq \bar{\theta}_n \leq \theta_0 + n^{-1/2}u^{\bullet}(\ell)$ (componentwise). Employing (4.59), we obtain $XII_2 = O(\sqrt{n})/N$ almost surely, where the $O(\sqrt{n})$ term does not depend on N . Choosing N large, we obtain $XII_2 = o(\sqrt{n})$ almost surely and we conclude that $XII = o_p(\sqrt{n})$. *Step 3* is completed.

Regarding *Step 4* we establish the following bound:

$$\begin{aligned} & \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(F(\tilde{\gamma}_t(u)x) - F(x) \right) - xf(x)\Omega'u \right| \\ & \leq \underbrace{\sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right) \right|}_{=XIII} \\ & \quad + \underbrace{\sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| xf(x) \frac{1}{n} \sum_{t=1}^n D'_t u - xf(x)\Omega'u \right|}_{=XIV} \\ & \quad + \underbrace{\sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(F(x\gamma_t(u)) - F(x) \right) - xf(x) \frac{1}{n} \sum_{t=1}^n D'_t u \right|}_{=XV}, \end{aligned} \tag{4.69}$$

where $XIII = O_p(n^{-1/2})$ by (4.60). Further, (4.34) and the ergodic theorem imply

$$XIV \leq \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} |x|f(x) \left\| \frac{1}{n} \sum_{t=1}^n D_t - \Omega \right\| \left\| u \right\| \leq A|\underline{x}|M \left\| \frac{1}{n} \sum_{t=1}^n D_t - \Omega \right\| \xrightarrow{a.s.} 0.$$

Regarding the last term, we use the mean value theorem and (4.34) to obtain

$$\begin{aligned} XV &= \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(xf(x\tilde{\gamma}_t)(\gamma_t(u) - 1) - xf(x) \frac{1}{\sqrt{n}} D'_t u \right) \right| \\ &\leq \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(xf(x)(\gamma_t(u) - 1) - xf(x) \frac{1}{\sqrt{n}} D'_t u \right) \right| \\ &\quad + \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(xf(x\tilde{\gamma}_t)(\gamma_t(u) - 1) - xf(x)(\gamma_t(u) - 1) \right) \right| \end{aligned}$$

$$\begin{aligned}
 & \leq \underbrace{\frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \left| (\gamma_t(u) - 1) - \frac{1}{\sqrt{n}} D'_t u \right|}_{XV_1} \\
 & \quad + \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n x(f(x\bar{\gamma}_t) - f(x))(\gamma_t(u) - 1) \right|}_{XV_2}
 \end{aligned}$$

with $\bar{\gamma}_t$ being between $\gamma_t(u)$ and 1. For n sufficiently large such that $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$, a second-order Taylor expansion gives

$$\begin{aligned}
 XV_1 &= \frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \frac{1}{\sigma_t(\theta_0)} \left| \sigma_t(\theta_0 + n^{-1/2}u) - \sigma_t(\theta_0) - \frac{1}{\sqrt{n}} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'} u \right| \\
 &= \frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \frac{1}{\sigma_t(\theta_0)} \left| \frac{1}{2n} u' \frac{\partial^2 \sigma_t(\bar{\theta}_n)}{\partial \theta \partial \theta'} u \right| \\
 &\leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} \left\| \frac{1}{\sigma_t(\bar{\theta}_n)} \frac{\partial^2 \sigma_t(\bar{\theta}_n)}{\partial \theta \partial \theta'} \right\| \\
 &\leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n \sup_{\sqrt{n}\|\theta - \theta_0\| \leq A} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\sqrt{n}\|\theta - \theta_0\| \leq A} \|H_t(\theta)\| \leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n T_t V_t
 \end{aligned}$$

with $\bar{\theta}_n$ being between θ_0 and $\theta_0 + n^{-1/2}u$. The Cauchy-Schwarz inequality, the uniform ergodic theorem and Assumption 4.9 yield

$$\frac{1}{n} \sum_{t=1}^n T_t V_t \leq \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[T_t^2] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n V_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[V_t^2] < \infty}$$

and we conclude that $XV_1 = O(n^{-1/2})$ almost surely. Before turning to XV_2 , we establish two auxiliary results:

- (i) $\frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| = O(1)$ almost surely;
- (ii) $\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} |f(x\bar{\gamma}_t) - f(x)| = o_p(1)$.

Statement (i) follows from (4.39) and (4.59) as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| \leq \frac{A}{n} \sum_{t=1}^n T_t U_t \leq A \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[T_t^2] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^2 \right)^{\frac{1}{2}}}_{\stackrel{a.s.}{\rightarrow} \mathbb{E}[U_t^2] < \infty}.$$

To show (ii), we note that the Cauchy-Schwarz inequality and Assumption 4.9 yield $\mathbb{E}[(T_t U_t)^3] \leq \mathbb{E}[T_t^6]^{\frac{1}{2}} \mathbb{E}[U_t^6]^{\frac{1}{2}} < \infty$. For every $\varepsilon > 0$ and for n sufficiently large such that $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$, we have

$$\begin{aligned} & \mathbb{P} \left[\sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| \geq \varepsilon \right] \leq \mathbb{P} \left[A \max_{1 \leq t \leq n} T_t U_t \geq \varepsilon \sqrt{n} \right] \\ & \leq \mathbb{P} \left[A^3 \max_{1 \leq t \leq n} (T_t U_t)^3 \geq \varepsilon^3 n^{3/2} \right] \leq \frac{A^3}{n^{3/2} \varepsilon^3} \mathbb{E} \left[\max_{1 \leq t \leq n} (T_t U_t)^3 \right] \\ & \leq \frac{A^3}{\sqrt{n} \varepsilon^3} \mathbb{E}[(T_t U_t)^3], \end{aligned}$$

which converges to 0, and thus we obtain $\sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| = o_p(1)$. Because $\bar{\gamma}_t$ lies between $\gamma_t(u)$ and 1, it follows that $\sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\bar{\gamma}_t - 1| = o_p(1)$. Thus, for sufficiently large n , we have $x\bar{\gamma}_t \in \mathcal{X}_\delta$ with probability close to one. Then, statement (ii) follows from the fact that f is uniformly continuous on \mathcal{X}_δ . Employing both auxiliary results, we obtain

$$\begin{aligned} XV_2 & \leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \frac{1}{\sqrt{n}} \sum_{t=1}^n |x| |f(x\bar{\gamma}_t) - f(x)| |\gamma_t(u) - 1| \\ & \leq |x| \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} |f(x\bar{\gamma}_t) - f(x)| \frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| = o_p(1). \end{aligned}$$

Thus XV is $o_p(1)$, which completes *Step 4*.

Concerning *Step 5* we obtain for each $\varepsilon > 0$

$$\begin{aligned} & \mathbb{P} \left[\sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}} - x f(x) \Omega' \sqrt{n} (\hat{\theta}_n - \theta_0) \right| \geq \varepsilon \right] \\ & \leq \mathbb{P} \left[\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \bar{\gamma}_t(u)x\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}} - x f(x) \Omega' u \right| \geq \varepsilon \right] \\ & \quad + \mathbb{P} \left[\sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right] \\ & \leq \mathbb{P} \left[\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(F(\bar{\gamma}_t(u)x) - F(x) \right) - x f(x) \Omega' u \right| \geq \frac{\varepsilon}{2} \right] \\ & \quad + \mathbb{P} \left[\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)/\sqrt{n}| \geq \frac{\varepsilon}{2} \right] + \mathbb{P} \left[\sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right]. \end{aligned}$$

Since $\sqrt{n} \|\hat{\theta}_n - \theta_0\| = O_p(1)$ by Theorem 4.2, the third term can be made arbitrarily small for large n by choosing A sufficiently large. Given A , the first two terms

converge to zero by *Step 3* and *Step 4*, which completes *Step 5*.

Regarding *Step 6* we refer to Bahadur (1966, Lemma 1). Replacing ξ by ξ_α in the proof and choosing the sequences a_n and b_n to satisfy $a_n \sim n^{-\varrho} \log n$ and $b_n \sim n^\psi$ as $n \rightarrow \infty$, where $\psi = (1 - \varrho)/2$, it follows that

$$\mathbb{H}_{n,\alpha} = \sup_{x \in \mathcal{I}_n} \left| (\mathbb{F}_n(x) - \mathbb{F}_n(\xi_\alpha)) - (F(x) - F(\xi_\alpha)) \right| = O(n^{-(\varrho+\psi)} \log n)$$

almost surely as $n \rightarrow \infty$. Inserting the definition of ψ and inflating the term by \sqrt{n} leads to $\sqrt{n} \mathbb{H}_{n,\alpha} = O(n^{-\varrho/2} \log n)$ almost surely as $n \rightarrow \infty$. Together with the triangle inequality, we establish

$$\sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \leq 2\sqrt{n} \mathbb{H}_{n,\alpha} = O(n^{-\varrho/2} \log n),$$

which completes *Step 6*.

Regarding *Step 7* we bound

$$\begin{aligned} & \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & \leq 2 \sup_{x \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \mathbb{F}_n(x)) - xf(x)\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \\ & \quad + \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & \quad + \sup_{x,y \in \mathcal{I}_n} \left| (xf(x) - yf(y))\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right|. \end{aligned}$$

Taking n sufficiently large such that $\mathcal{I}_n \subset \mathcal{X}$, the first term on the right-hand side vanishes in probability by *Step 5*. The second term vanishes almost surely by *Step 6*. The last term can be bounded as follows:

$$\begin{aligned} & \sup_{x,y \in \mathcal{I}_n} \left| (xf(x) - yf(y))\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \\ & \leq \sup_{x,y \in \mathcal{I}_n} |xf(x) - yf(y)| \|\Omega\| \sqrt{n} \|\hat{\theta}_n - \theta_0\|. \end{aligned}$$

Since $f(x)$, and hence $xf(x)$, is continuous in a neighborhood of ξ_α by Assumption 4.5(ii) and \mathcal{I}_n shrinks to ξ_α we have $\sup_{x,y \in \mathcal{I}_n} |xf(x) - yf(y)| \rightarrow 0$. Together with $\sqrt{n} \|\hat{\theta}_n - \theta_0\| = O_p(1)$ (Theorem 4.2) we find that the last term converges in probability to 0, which completes *Step 7*.

To verify that replacing any $\hat{\mathbb{F}}_n(\cdot)$ by $\hat{\mathbb{F}}_n(\cdot -)$ does not alter the result, we note

that $\hat{\mathbb{F}}_n(x - n^{-1}) \leq \hat{\mathbb{F}}_n(x-) \leq \hat{\mathbb{F}}_n(x) \leq \hat{\mathbb{F}}_n(x + n^{-1})$ for all $x \in \mathcal{I}_n$ (similarly for y). Setting $\bar{\mathcal{I}}_n = (\xi_\alpha - \bar{a}_n, \xi_\alpha + \bar{a}_n)$ with $\bar{a}_n = a_n + n^{-1}$, we can bound $\sup_{x,y \in \mathcal{I}_n} |\sqrt{n}(\hat{\mathbb{F}}_n(x-) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y))|$ and $\sup_{x,y \in \mathcal{I}_n} |\sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y))|$ by

$$\begin{aligned} & \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & \quad + 2 \sup_{y \in \mathcal{I}_n} \sqrt{n} \left(F(y + n^{-1}) - F(y - n^{-1}) \right). \end{aligned} \quad (4.70)$$

The first term in (4.70) vanishes in probability by *Step 7* as $\bar{a}_n \sim a_n$. Regarding the second term, the mean value theorem implies

$$2 \sup_{y \in \mathcal{I}_n} \sqrt{n} \left(F\left(y + \frac{1}{n}\right) - F\left(y - \frac{1}{n}\right) \right) = \frac{4}{\sqrt{n}} \sup_{y \in \mathcal{I}_n} f(y + \varepsilon_n),$$

where $|\varepsilon_n| \leq n^{-1}$. Since $\frac{4}{\sqrt{n}} \rightarrow 0$ and $\sup_{y \in \mathcal{I}_n} f(y + \varepsilon_n) \rightarrow f(\xi_\alpha)$ the term vanishes, which completes the proof. \square

Remark 4.4. *Step 5* is closely related to Lemma 3.2 of Gao and Song (2008) with Ω corresponding to their $\mathbf{e}/2$. Whereas in *Step 5* we establish the uniformity over a compact neighborhood of ξ_α , they claim –without formal proof– uniform convergence in probability over \mathbb{R} assuming differentiability of f and $\sup_{x \in \mathbb{R}} x^2 |f'(x)| < \infty$.

4.A.2 Bootstrap Lemmas

Henceforth we use \mathbb{P}^* , \mathbb{E}^* , Var^* and Cov^* to denote the probability, expectation, variance and covariance conditional on \mathcal{F}_n .

Lemma 4.4. *Suppose Assumptions 4.1–4.3, 4.4(i), 4.5(i) and 4.5(iii) hold.*

- (i) *If in addition Assumption 4.9(i) holds with $a = 4$, then $\mathbb{E}^*[\eta_t^{*m}] \xrightarrow{a.s.} \mathbb{E}[\eta_t^m]$ for $m \in \{1, 2, 3, 4\}$.*
- (ii) *If in addition Assumptions 4.6, 4.7 and 4.9(i) hold with $a = -1, 4$, then we have $\mathbb{E}^*[\eta_t^{*m} \mathbb{1}_{\{\eta_t^* < \xi_{n,\alpha}\}}] \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]$ for $m \in \{0, 1, 2, 3, 4\}$.*

Proof. Lemma 4.2 gives $\mathbb{E}^*[\eta_t^{*m} \mathbb{1}_{\{\eta_t^* < u\}}] = \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{\hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < u\}}]$. Taking $u = \infty$ proves the first claim, whereas the second claim follows from $\mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < u\}}]$ being continuous in u and $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$ by Theorem 4.1. \square

Lemma 4.5. *Suppose Assumptions 4.1–4.3, 4.4(i), 4.5(i), 4.5(iii), 4.6 and 4.9(i)–(ii) hold with $a = \pm 4$. Then, we have $\hat{\theta}_n^* \xrightarrow{p^*} \theta_0$ almost surely.*

Proof. Let $\nu > 0$ and set $\mathcal{B} = \{\theta \in \Theta : \|\theta - \theta_0\| \geq \nu\}$; We establish the result in the following three steps:

Step 1: we obtain $L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + R_n^*(\theta)$ with $\sup_{\theta \in \Theta} |R_n^*(\theta)| \xrightarrow{p^*} 0$ almost surely;

Step 2: There exists a $\zeta < 0$ such that $\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) < \zeta/2 + S_n^*$ with $S_n^* \xrightarrow{p^*} 0$ almost surely;

Step 3: we show $\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \xrightarrow{a.s.} 0$.

Regarding *Step 1* we find

$$L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left\{ \eta_t^{*2} - \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \eta_t^{*2} + \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right\},$$

where $\frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \xrightarrow{p^*} 1$ almost surely since

$$\mathbb{E}^* \left[\frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \right] = \mathbb{E}^* [\eta_t^{*2}] \xrightarrow{a.s.} 1 \quad \text{and} \quad \mathbb{V}\text{ar}^* \left[\frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \right] = \frac{1}{n} \mathbb{V}\text{ar}^* [\eta_t^{*2}] \xrightarrow{a.s.} 0$$

by Lemma 4.4. It remains to show the negligibility of the initial conditions, i.e.

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left\{ \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right\} \right| \xrightarrow{a.s.} 0 \quad (4.71)$$

and

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right) \eta_t^{*2} \right| \xrightarrow{p^*} 0 \quad (4.72)$$

almost surely. The inequality $\log(1+x) \leq x$ for all $x > -1$ and Assumption 4.4(i) yield

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left(\log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right) \right|$$

$$\begin{aligned}
 &= \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left(\log \frac{\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} - \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta}_n)} \right) \right| \\
 &\leq \sup_{\theta \in \Theta} \frac{2}{n} \sum_{t=1}^n \left| \log \frac{\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} \right| = \sup_{\theta \in \Theta} \frac{4}{n} \sum_{t=1}^n \left| \log \frac{\tilde{\sigma}_t(\theta)}{\sigma_t(\theta)} \right| \\
 &= \sup_{\theta \in \Theta} \frac{4}{n} \sum_{t=1}^n \left| \log \left(1 + \frac{\tilde{\sigma}_t(\theta) - \sigma_t(\theta)}{\sigma_t(\theta)} \right) \right| \\
 &\leq \frac{4}{n} \sum_{t=1}^n \log \left(1 + \frac{C_1 \rho^t}{\underline{\omega}} \right) \leq \frac{4}{n} \sum_{t=1}^n \frac{C_1 \rho^t}{\underline{\omega}} \leq \frac{4C_1}{\underline{\omega}(1-\rho)n} \xrightarrow{a.s.} 0
 \end{aligned}$$

verifying (4.71). Further, Assumption 4.4(i) and (4.29) imply

$$\begin{aligned}
 &\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) \eta_t^{*2} \right| \leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right| \eta_t^{*2} \\
 &= \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta}_n)} + \frac{\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} \right| \eta_t^{*2} \\
 &\leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left(\frac{|\tilde{\sigma}_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta}_n)|}{\sigma_t^2(\hat{\theta}_n)} + \frac{|\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)|}{\sigma_t^2(\theta)} \right) \eta_t^{*2} \\
 &\leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left(\frac{|\tilde{\sigma}_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)|^2}{\sigma_t^2(\hat{\theta}_n)} + 2 \frac{|\tilde{\sigma}_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)|}{\sigma_t(\hat{\theta}_n)} \right. \\
 &\quad \left. + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|^2}{\sigma_t^2(\theta)} + 2 \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\sigma_t(\theta)} \right) \eta_t^{*2} \\
 &\leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\underline{\omega}^2} \left(\frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} + 2 \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} + 2 \frac{C_1 \rho^t}{\underline{\omega}} \right) \eta_t^{*2} \\
 &\leq \left(\frac{2C_1^2}{\underline{\omega}^4} + \frac{4C_1}{\underline{\omega}^3} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2}.
 \end{aligned}$$

To verify (4.72) are left to show that $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2} \xrightarrow{p^*} 0$ almost surely. For every $\varepsilon > 0$, Markov's inequality implies

$$\mathbb{P}^* \left[\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2} \geq \varepsilon \right] \leq \frac{1}{\varepsilon} \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \mathbb{E}^* [\eta_t^{*2}]$$

As $\mathbb{E}^* [\eta_t^{*2}] \xrightarrow{a.s.} 1$ (Lemma 4.4), it remains to show that $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \xrightarrow{a.s.} 0$. We

have

$$\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\theta_0) \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta_0)} \leq \left(\frac{1}{n} \sum_{t=1}^n \rho^{2t} \sigma_t^4(\theta_0) \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^4(\hat{\theta}_n)}{\sigma_t^4(\theta_0)} \right)^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ (Theorem 4.1) such that $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ almost surely, the uniform ergodic theorem and Assumption 4.9(i) result in

$$\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^4(\hat{\theta}_n)}{\sigma_t^4(\theta_0)} \stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n T_t^4 \xrightarrow{a.s.} \mathbb{E}[T_t^4] < \infty.$$

In addition, we have for $\delta > 0$

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^{2t} \sigma_t^4(\theta_0) > \delta] \leq \sum_{t=1}^{\infty} \frac{\rho^{st/2} \mathbb{E}[\sigma_t^s(\theta_0)]}{\delta^{s/(4)}} = \frac{\mathbb{E}[\sigma_t^s(\theta_0)]}{\delta^{s/(4)}(1 - \rho^{s/2})} < \infty$$

such that the Borel-Cantelli Lemma implies $\rho^{2t} \sigma_t^4(\theta_0) \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$. Therefore, $\frac{1}{n} \sum_{t=1}^n \rho^{2t} \sigma_t^4(\theta_0) \xrightarrow{a.s.} 0$ follows by Cesàro's lemma. Combining results, we establish $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \xrightarrow{a.s.} 0$, which verifies (4.72) and completes *Step 1*.

Consider *Step 2*; by compactness of \mathcal{B} the Heine-Borel theorem entails that there exists a finite number of neighborhoods of size smaller than $1/k$, i.e. $\mathcal{V}_k(\theta_1), \dots, \mathcal{V}_k(\theta_K)$ with $K = K(k) \in \mathbb{N}$, covering \mathcal{B} . We have

$$\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \max_{i=1, \dots, K} \sup_{\theta \in \mathcal{V}_k(\theta_i) \cap \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n)$$

Next, we fix $i \in \{1, \dots, K\}$. With regard to *Step 1*, we obtain for each $M > 1$

$$\begin{aligned} & L_n^*(\theta) - L_n^*(\hat{\theta}_n) \\ &= \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} > M \right\}} \underbrace{\left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right)}_{\geq 0} \\ & \quad + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M \right\}} \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + R_n^*(\theta) \\ & \leq \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} > M \right\}} \left(1 + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M \right\}} \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} (1 - \eta_t^{*2}) \\ & \quad + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M \right\}} \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + R_n^*(\theta) \end{aligned}$$

such that

$$\begin{aligned}
 & \sup_{\theta \in \mathcal{V}_k(\theta_i) \cap \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) \\
 & \leq \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} > M\right\}} \left(1 + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)}\right)}_I \\
 & \quad + \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M\right\}} \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} (1 - \eta_t^{*2})}_{II} \\
 & \quad + \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M\right\}} \left(1 - \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)}\right)}_{III} + \underbrace{\sup_{\theta \in \Theta} |R_n^*(\theta)|}_{IV}.
 \end{aligned}$$

Subsequently, we consider each term in turn. Regarding I , take k sufficiently large such that $\dot{\theta}$ satisfying $\|\dot{\theta} - \theta_0\| \leq 1/k$ yields $\dot{\theta} \in \mathcal{V}(\theta_0)$. The uniform ergodic theorem, the inequality $\log(x) \leq x$ for all $x > 0$ and the Cauchy-Schwarz inequality imply

$$\begin{aligned}
 I & \xrightarrow{a.s.} \mathbb{E} \left[\sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} > M\right\}} \left(1 + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)}\right) \right] \\
 & \leq \mathbb{E} \left[\mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left(1 + \log \frac{\sigma_t^2 T_t^2}{\underline{\omega}^2}\right) \right] \\
 & = \mathbb{E} \left[\mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left(1 - 2 \log \underline{\omega} + \frac{4}{s} \log \sigma_t^{s/2} + 2 \log T_t\right) \right] \\
 & \leq \mathbb{E} \left[\mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left(1 - 2 \log \underline{\omega} + \frac{4}{s} \sigma_t^{s/2} + 2 T_t\right) \right] \\
 & \leq \underbrace{\left(\mathbb{E} \left[\left(1 - 2 \log \underline{\omega} + \frac{4}{s} \sigma_t^{s/2} + 2 T_t\right)^2 \right] \right)^{\frac{1}{2}}}_{I_1} \underbrace{\left(\mathbb{P} \left[\sigma_t^2 T_t^2 > M \underline{\omega}^2 \right] \right)^{\frac{1}{2}}}_{I_2}
 \end{aligned}$$

with $\sigma_t = \sigma_t(\theta_0)$. Employing (4.36) we find that

$$I_1 \leq 4 \left(1 + (2 \log \underline{\omega})^2 + \frac{16}{s^2} \mathbb{E}[\sigma_t^s] + 4 \mathbb{E}[T_t^2] \right) < \infty$$

and using Markov's inequality the second subterm can be bounded by

$$I_2 \leq \mathbb{P}\left[T_t^2 > M\underline{\omega}^2/2\right] + \mathbb{P}\left[\sigma_t^2 > M\underline{\omega}^2/2\right] \leq \frac{2}{M\underline{\omega}^2} \mathbb{E}[T_t^2] + \left(\frac{2}{\sqrt{M\underline{\omega}}}\right)^s \mathbb{E}[\sigma_t^s].$$

Since I_1 can be made arbitrarily small by the choice of M we get $I = o(1)$ almost surely. Further, for given M , Lemma 4.4 entails

$$\left|\mathbb{E}^*[II]\right| \leq M \left|1 - \mathbb{E}^*[\eta_t^{*2}]\right| \xrightarrow{a.s.} 0 \quad \text{and} \quad \mathbb{V}\text{ar}^*[II] \leq \frac{M^2}{n} \mathbb{V}\text{ar}^*[\eta_t^{*2}] \xrightarrow{a.s.} 0$$

such that $II \xrightarrow{p^*} 0$ almost surely. Consider III ; the uniform ergodic theorem yields

$$III \xrightarrow{a.s.} \mathbb{E} \left[\sup_{\substack{||\dot{\theta} - \theta_0|| \leq 1/k \\ ||\theta - \theta_i|| \leq 1/k}} \mathbb{1}_{\left\{ \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M \right\}} \left(1 - \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \right) \right]$$

and the right-hand side approaches

$$\mathbb{E} \left[1 - \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} + \log \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} \right] \quad (4.73)$$

as M and k grow large. Thus, almost surely, III can be made arbitrarily close to (4.73) by choosing M and k sufficiently large. Further, since $\theta_i \in \mathcal{B}$, we have $\theta_i \neq \theta_0$ and Assumption 4.3 implies $\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} \neq 1$ almost surely. The elementary inequality $1 - x + \log x \leq 0$ for $x > 0$, which holds with equality if and only if $x = 1$, implies that (4.73) is strictly smaller than 0. We conclude that there exists a $\zeta_i < 0$ such that $III < \zeta_i$ holds for sufficiently large M and k and n almost surely. Set $\zeta = \max_{i=1, \dots, K} \zeta_i$, which satisfies $\zeta < 0$. Combining results we complete *Step 2*.

Consider *Step 3*; if $\hat{\theta}_n^* \in \mathcal{B}$, then (4.15) yields

$$\sup_{\theta \in \mathcal{B}} L_n^*(\theta) = L_n^*(\hat{\theta}_n^*) \geq L_n^*(\hat{\theta}_n).$$

and by monotonicity of the probability measure \mathbb{P}^* we obtain

$$\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \leq \mathbb{P}^* \left[\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) \geq 0 \right].$$

Together with *Step 2* we obtain

$$\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \leq \mathbb{P}^*[\zeta/2 + S_n^* > 0] + o(1) \leq \mathbb{P}^*[|S_n^*| > \zeta/2] + o(1) = o(1)$$

almost surely, which completes *Step 3* and establishes the lemma's claim. \square

Lemma 4.6. *Suppose Assumptions 4.1–4.4, 4.5(i), 4.5(iii), 4.6 and 4.9 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have $\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \xrightarrow{P^*} -2J$ almost surely for $\check{\theta}_n$ between $\hat{\theta}_n^*$ and $\hat{\theta}_n$.*

Proof. We have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) &= \underbrace{\frac{1}{n} \sum_{t=1}^n \left(\frac{\epsilon_t^{*2}}{\tilde{\sigma}_t^2(\check{\theta}_n)} - 1 \right) \tilde{H}_t(\check{\theta}_n)}_I \\ &\quad - \underbrace{\frac{1}{n} \sum_{t=1}^n \left(3 \frac{\epsilon_t^{*2}}{\tilde{\sigma}_t^2(\check{\theta}_n)} - 1 \right) \tilde{D}_t(\check{\theta}_n) \tilde{D}_t'(\check{\theta}_n)}_{II}. \end{aligned}$$

Employing $\epsilon_t^* = \tilde{\sigma}_t(\hat{\theta}_n) \eta_t^*$ the first term can be expanded as follows:

$$\begin{aligned} I &= \underbrace{\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \eta_t^{*2}}_{I_1} + \underbrace{\frac{1}{n} \sum_{t=1}^n \left(\frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{H}_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \right) \eta_t^{*2}}_{I_2} \\ &\quad - \underbrace{\frac{1}{n} \sum_{t=1}^n \tilde{H}_t(\check{\theta}_n)}_{I_3}. \end{aligned}$$

Consider I_1 ; we take $\varepsilon > 0$ and denote the unit vectors spanning \mathbb{R}^r by e_1, \dots, e_r . Since $\frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} H_t(\theta_2)$ is continuous in θ_1 and θ_2 we can take $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

$$\begin{aligned} \mathbb{E}[e_i' H_t e_j] - \varepsilon &< \mathbb{E} \left[\inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \right] \\ &\leq \mathbb{E} \left[\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \right] < \mathbb{E}[e_i' H_t e_j] + \varepsilon \end{aligned}$$

for all $i, j = 1, \dots, r$. Since $\check{\theta}_n$ lies between $\hat{\theta}_n^*$ and $\hat{\theta}_n$, Theorem 4.1 and Lemma 4.5 imply $\check{\theta}_n \xrightarrow{P^*} \theta_0$ almost surely. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ and $\check{\theta}_n \xrightarrow{P^*} \theta_0$ almost surely, we

have $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ almost surely and $\check{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ with conditional probability close to one almost surely. In such case, we have for all pairs (i, j)

$$L_n^*(i, j) \leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e_i' H_t(\check{\theta}_n) e_j \eta_t^{*2} \leq U_n^*(i, j)$$

with

$$L_n^*(i, j) = \frac{1}{n} \sum_{t=1}^n \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \eta_t^{*2}$$

$$U_n^*(i, j) = \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \eta_t^{*2}.$$

Using the uniform ergodic theorem, the conditional mean of the upper bound satisfies

$$\mathbb{E}^*[U_n^*(i, j)] = \mathbb{E}^*[\eta_t^{*2}] \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j$$

$$\xrightarrow{a.s.} \mathbb{E} \left[\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \right] < \mathbb{E}[e_i' H_t e_j] + \varepsilon.$$

whereas its conditional variance vanishes:

$$\text{Var}^*[U_n^*(i, j)] = \text{Var}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n \left(\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \right)^2$$

$$\leq \text{Var}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n S_t^4 T_t^4 V_t^2$$

$$\leq \text{Var}^*[\eta_t^{*2}] \frac{1}{n} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n S_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[S_t^{12}] < \infty}^{\frac{1}{3}} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[T_t^{12}] < \infty}^{\frac{1}{3}} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n V_t^6 \right)}_{\xrightarrow{a.s.} \mathbb{E}[V_t^6] < \infty}^{\frac{1}{3}} \xrightarrow{a.s.} 0.$$

Similarly, we obtain for the lower bound

$$\mathbb{E}^*[L_n^*(i, j)] \xrightarrow{a.s.} \mathbb{E} \left[\inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e_i' H_t(\theta_2) e_j \right] > \mathbb{E}[e_i' H_t e_j] - \varepsilon$$

and $\mathbb{V}\text{ar}^*[L_n^*(i, j)] \xrightarrow{a.s.} 0$. Taking $\varepsilon \searrow 0$ subsequently, we get

$$\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e_i' H_t(\check{\theta}_n) e_j' \eta_t^{*2} \xrightarrow{p^*} \mathbb{E}[e_i' H_t e_j]$$

almost surely for all pairs (i, j) , which in turn yields $I_1 \xrightarrow{p^*} \mathbb{E}[H_t]$ almost surely. Regarding I_2 , for any $\theta_1, \theta_2 \in \Theta$ Assumption 4.4(i) implies

$$\begin{aligned} \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right| &= \left| \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\tilde{\sigma}_t(\theta_2)} \right| \\ &\leq \frac{|\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)|}{\tilde{\sigma}_t(\theta_2)} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{|\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)|}{\tilde{\sigma}_t(\theta_2)} \\ &\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{C_1 \rho^t}{\underline{\omega}} = \frac{C_1 \rho^t}{\underline{\omega}} \left(1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \end{aligned} \quad (4.74)$$

and combined with the elementary inequalities (4.29) (with $m = 1$) yields

$$\begin{aligned} \left| \frac{\tilde{\sigma}_t^2(\theta_1)}{\tilde{\sigma}_t^2(\theta_2)} - \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} \right| &\leq \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right|^2 + 2 \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right| \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \\ &\leq \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} \left(1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right)^2 + \frac{2C_1 \rho^t}{\underline{\omega}} \left(1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \\ &\leq \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right)^2 \leq \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} \right). \end{aligned} \quad (4.75)$$

It follows that

$$\begin{aligned} \|I_2\| &\leq \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{H}_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \right\| \eta_t^{*2} \\ &= \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left(\tilde{H}_t(\check{\theta}_n) - H_t(\check{\theta}_n) \right) + \left(\frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) H_t(\check{\theta}_n) \right\| \eta_t^{*2} \\ &\leq \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left\| \tilde{H}_t(\check{\theta}_n) - H_t(\check{\theta}_n) \right\| + \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right| \|H_t(\check{\theta}_n)\| \right\} \eta_t^{*2} \\ &\leq \frac{1}{n} \sum_{t=1}^n \left\{ \left(\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} + \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \right) \frac{C_1 \rho^t}{\underline{\omega}} \left(1 + \|H_t(\check{\theta}_n)\| \right) \right. \\ &\quad \left. + \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \|H_t(\check{\theta}_n)\| \right\} \eta_t^{*2} \\ &\leq \left(\frac{5C_1}{\underline{\omega}} + \frac{6C_1^2}{\underline{\omega}^2} + \frac{2C_1^3}{\underline{\omega}^3} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left(1 + \|H_t(\check{\theta}_n)\| \right) \eta_t^{*2}, \end{aligned}$$

where the third inequality comes from (4.32) and (4.75). In the case of $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ and $\check{\theta}_n \in \mathcal{V}(\theta_0)$, we get

$$\frac{1}{n} \sum_{t=1}^n \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left(1 + \|H_t(\check{\theta}_n)\| \right) \eta_t^{*2} \leq \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2}.$$

For any $\delta > 0$ we find

$$\mathbb{P}^* \left[\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2} \geq \delta \right] = \frac{\mathbb{E}^*[\eta_t^{*2}]}{\delta} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t).$$

using Markov's inequality. Moreover, for $\varepsilon > 0$ we have

$$\begin{aligned} \sum_{t=1}^{\infty} \mathbb{P} \left[\rho^t (1 + S_t^2 T_t^2) (1 + V_t) > \varepsilon \right] &\leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + V_t)]}{\varepsilon} \\ &= \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + V_t)]}{\varepsilon(1 - \rho)} < \infty \end{aligned}$$

such that the Borel-Cantelli Lemma implies $\rho^t (1 + S_t^2 T_t^2) (1 + V_t) \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$. Therefore, $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \xrightarrow{a.s.} 0$ follows by Césaro's lemma and we get $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2} \xrightarrow{p^*} 0$ almost surely. Combining results gives $\|I_2\| \xrightarrow{p^*} 0$ almost surely. Similar to the proof of Lemma 4.2(iii), we establish $I_3 \xrightarrow{p^*} \mathbb{E}[H_t]$ almost surely using $\check{\theta}_n \xrightarrow{p^*} \theta_0$ almost surely. Combining results we establish that $I = I_1 + I_2 - I_3 \xrightarrow{p^*} 0$ almost surely. Consider the second term and expand

$$\begin{aligned} II = & 3 \underbrace{\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \eta_t^{*2}}_{II_1} \\ & + 3 \underbrace{\frac{1}{n} \sum_{t=1}^n \left(\frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right) \eta_t^{*2}}_{II_2} \\ & - \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\check{\theta}_n) D'_t(\check{\theta}_n)}_{II_3}. \end{aligned}$$

We treat the subterms of II analogously to the subterms of I . We begin with II_1 and take $\varepsilon > 0$. Since $\frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} D_t(\theta_2) D'_t(\theta_2)$ is continuous in θ_1 and θ_2 we can take

$\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

$$\begin{aligned} \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon &< \mathbb{E} \left[\inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] \\ &\leq \mathbb{E} \left[\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

for all $i, j = 1, \dots, r$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ and $\check{\theta}_n \xrightarrow{P^*} \theta_0$ almost surely, we have $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ almost surely and $\check{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$ with conditional probability close to one almost surely. In such case, we have for all pairs (i, j)

$$\bar{L}_n^*(i, j) \leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i D_t(\check{\theta}_n) D'_t(\check{\theta}_n) e_j \eta_t^{*2} \leq \bar{U}_n^*(i, j)$$

with

$$\begin{aligned} \bar{L}_n^*(i, j) &= \frac{1}{n} \sum_{t=1}^n \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \eta_t^{*2} \\ \bar{U}_n^*(i, j) &= \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \eta_t^{*2}. \end{aligned}$$

Using the uniform ergodic theorem, the conditional mean of the upper bound satisfies

$$\begin{aligned} \mathbb{E}^*[\bar{U}_n^*(i, j)] &= \mathbb{E}^*[\eta_t^{*2}] \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \\ &\xrightarrow{a.s.} \mathbb{E} \left[\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

whereas its conditional variance vanishes:

$$\begin{aligned} \mathbb{V}\text{ar}^*[\bar{U}_n^*(i, j)] &= \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n \left(\sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right)^2 \\ &\leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n S_t^4 T_t^4 U_t^4 \\ &\leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n S_t^{12} \right)^{\frac{1}{3}}}_{\xrightarrow{a.s.} \mathbb{E}[S_t^{12}] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^{12} \right)^{\frac{1}{3}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^{12}] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^{12} \right)^{\frac{1}{3}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^{12}] < \infty} \xrightarrow{a.s.} 0. \end{aligned}$$

Similarly, we obtain for the lower bound

$$\mathbb{E}^*[\bar{L}_n^*(i, j)] \xrightarrow{a.s.} \mathbb{E}\left[\inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j\right] > \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon$$

and $\mathbb{V}\text{ar}^*[\bar{L}_n^*(i, j)] \xrightarrow{a.s.} 0$. Next, we take $\varepsilon \searrow 0$ and get

$$\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i D_t(\check{\theta}_n) D'_t(\check{\theta}_n) e'_j \eta_t^{*2} \xrightarrow{p^*} \mathbb{E}[e'_i D_t D'_t e_j]$$

almost surely for all pairs (i, j) , which in turn yields $II_1 \xrightarrow{p^*} \mathbb{E}[D_t D'_t] = J$ almost surely. Regarding II_2 , we find

$$\begin{aligned} \|II_2\| &\leq \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| \eta_t^{*2} \\ &= \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left(\tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right) \right. \\ &\quad \left. + \left(\frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| \eta_t^{*2} \\ &\leq \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left\| \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| \right. \\ &\quad \left. + \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right| \|D_t(\check{\theta}_n)\|^2 \right\} \eta_t^{*2} \\ &\leq \frac{1}{n} \sum_{t=1}^n \left\{ \left(\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} + \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \right) \right. \\ &\quad \times \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left(1 + \|D_t(\check{\theta}_n)\|^2 \right) \\ &\quad \left. + \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \|D_t(\check{\theta}_n)\|^2 \right\} \eta_t^{*2} \\ &\leq \left(\frac{6C_1}{\underline{\omega}} + \frac{11C_1^2}{\underline{\omega}^2} + \frac{8C_1^3}{\underline{\omega}^3} + \frac{2C_1^4}{\underline{\omega}^4} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left(1 + \|D_t(\check{\theta}_n)\|^2 \right) \eta_t^{*2}, \end{aligned}$$

where the third inequality follows from (4.30) and (4.75). In the case of $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ and $\check{\theta}_n \in \mathcal{V}(\theta_0)$, we get

$$\frac{1}{n} \sum_{t=1}^n \rho^t \left(1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left(1 + \|D_t(\check{\theta}_n)\|^2 \right) \eta_t^{*2} \leq \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2}.$$

For any $\delta > 0$ we find

$$\mathbb{P}^* \left[\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2} \geq \delta \right] = \frac{\mathbb{E}^*[\eta_t^{*2}]}{\delta} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2).$$

using Markov's inequality. Moreover, for $\varepsilon > 0$ we have

$$\begin{aligned} \sum_{t=1}^{\infty} \mathbb{P} \left[\rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) > \varepsilon \right] &\leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + U_t^2)]}{\varepsilon} \\ &= \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + U_t^2)]}{\varepsilon(1 - \rho)} < \infty \end{aligned}$$

such that the Borel-Cantelli Lemma implies $\rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$. Therefore, $\frac{1}{n^2} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \xrightarrow{a.s.} 0$ follows by Césaro's lemma and we get $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2} \xrightarrow{p^*} 0$ almost surely. Combining results gives $\|II_2\| \xrightarrow{p^*} 0$ almost surely. Similar to the proof of Lemma 4.2(ii), we establish $II_3 \xrightarrow{p^*} \mathbb{E}[D_t D_t'] = J$ almost surely using $\check{\theta}_n \xrightarrow{p^*} \theta_0$ almost surely. Combining results we find $II = 3II_1 + 3II_2 - I_3 \xrightarrow{p^*} 3J + 0 - J = 2J$ almost surely. In conclusion, we have

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) = I - II \xrightarrow{p^*} -2J$$

almost surely, which completes the proof. \square

Lemma 4.7. *Suppose Assumptions 4.1–4.4, 4.5(i), 4.5(iii), 4.6, 4.9 and 4.10 hold with $a = -1, 4$, $b = 4$ and $c = 2$. Then, we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - 1) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \end{pmatrix} \xrightarrow{d^*} N(0, \Upsilon_\alpha) \quad \text{with} \quad \Upsilon_\alpha = \begin{pmatrix} (\kappa - 1)J & p_\alpha \Omega \\ p_\alpha \Omega' & \alpha(1 - \alpha) \end{pmatrix}$$

almost surely.

Proof. Set $\alpha_n = \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}]$ and expand

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - 1) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \end{pmatrix} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}]) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha_n \end{pmatrix} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\mathbb{E}^*[\eta_t^{*2}] - 1) \\ \alpha_n - \alpha \end{pmatrix}. \end{aligned}$$

Consider the second term; with regard to Remark 4.1 we have $\mathbb{E}^*[\eta_t^{*2}] = 1$ whenever $\hat{\theta}_n \in \mathring{\Theta}$ under Assumption 4.10. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0 \in \mathring{\Theta}$ by Theorem 4.1 and Assumption 4.6, we have $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\mathbb{E}^*[\eta_t^{*2}] - 1) = 0$ for sufficiently large n almost surely. Further,

$$\alpha_n = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} \stackrel{a.s.}{=} \frac{\lfloor n\alpha \rfloor + 1}{n} = \alpha + O(n^{-1})$$

and hence $\frac{1}{\sqrt{n}} \sum_{t=1}^n (\alpha_n - \alpha) \xrightarrow{a.s.} 0$. Using the Cramér-Wold device it remains to show that for each $\lambda = (\lambda'_1, \lambda'_2)' \in \mathbb{R}^{r+1}$ with $\|\lambda\| \neq 0$

$$\sum_{t=1}^n \frac{1}{\sqrt{n}} \lambda' \underbrace{\left(\hat{D}_t(\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}]) \right)}_{Z_{n,t}} \xrightarrow{d^*} N(0, \lambda' \Upsilon_\alpha \lambda)$$

almost surely. By construction, we have $\mathbb{E}[Z_{n,t}^*] = 0$. Further, we obtain

$$s_n^2 = \sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2}] = \lambda' \begin{pmatrix} \text{Var}^*[\eta_t^{*2}] \hat{J}_n & \text{Cov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}_n \\ \text{Cov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}'_n & \text{Var}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \end{pmatrix} \lambda.$$

Lemma 4.2 states $\hat{J}_n \xrightarrow{a.s.} J$ and $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$. Employing Lemma 4.4 yields

$$\begin{aligned} \text{Var}^*[\eta_t^{*2}] &= \mathbb{E}^*[\eta_t^{*4}] - \left(\mathbb{E}[\eta_t^{*2}] \right)^2 \xrightarrow{a.s.} \kappa - 1, \\ \text{Var}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] &= \alpha_n(1 - \alpha_n) \xrightarrow{a.s.} \alpha(1 - \alpha), \\ \text{Cov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] &= \mathbb{E}^*[\eta_t^{*2} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \mathbb{E}^*[\eta_t^{*2}] \alpha_n \xrightarrow{a.s.} p_\alpha \end{aligned}$$

and $s_n^2 \xrightarrow{a.s.} \lambda' \Upsilon_\alpha \lambda$ follows. Next, we verify Lindeberg condition. For an arbitrary $\varepsilon > 0$

$$\begin{aligned} & \sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}}] \\ & \leq \underbrace{\sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|\eta_t^*| > C\}}]}_I + \underbrace{\sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}} \mathbb{1}_{\{|\eta_t^*| \leq C\}}]}_{II} \end{aligned}$$

holds, where $C > 0$. Employing the elementary inequalities

$$(x + y)^z \leq 2^z(x^z + y^z) \quad (4.76)$$

and $|x - y|^z \leq x^z + y^z$ for all $x, y, z \geq 0$ we find that

$$\begin{aligned} Z_{n,t}^{*2} &\leq \frac{4}{n} \left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}])^2 + \lambda_2^2 (\mathbb{1}_{\{\eta_t^* < \hat{\epsilon}_{n,\alpha}\}} - \alpha_n)^2 \right) \\ &\leq \frac{4}{n} \left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right). \end{aligned}$$

Hence, we obtain

$$\begin{aligned} I &\leq \frac{4}{n} \sum_{t=1}^n \mathbb{E}^* \left[\left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{|\eta_t^*| > C\}} \right] \\ &= 4 \left(\lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^*[\eta_t^{*4} \mathbb{1}_{\{|\eta_t^*| > C\}}] + (\lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^*[\eta_t^{*2}]^2 + \lambda_2^2) \mathbb{E}^*[\mathbb{1}_{\{|\eta_t^*| > C\}}] \right) \\ &\stackrel{a.s.}{\rightarrow} 4 \left(\lambda'_1 J \lambda_1 \mathbb{E}[\eta_t^4 \mathbb{1}_{\{|\eta_t| > C\}}] + (\lambda'_1 J \lambda_1 \mathbb{E}[\eta_t^2]^2 + \lambda_2^2) \mathbb{E}[\mathbb{1}_{\{|\eta_t| > C\}}] \right) \end{aligned}$$

and choosing C sufficiently large yields $I \stackrel{a.s.}{\rightarrow} 0$. Given a value of C , we have

$$\begin{aligned} II &\leq \frac{4}{n} \sum_{t=1}^n \mathbb{E}^* \left[\left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \right. \\ &\quad \times \mathbb{1}_{\{||\lambda_1||(\eta_t^{*2} + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t|| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon\}} \mathbb{1}_{\{|\eta_t^*| \leq C\}} \left. \right] \\ &\leq \frac{4}{n} \sum_{t=1}^n \left((\lambda'_1 \hat{D}_t)^2 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \\ &\quad \times \mathbb{1}_{\{||\lambda_1|| (C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t|| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon\}} \\ &= 4 \left(\lambda'_1 \hat{J}_n \lambda_1 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{||\lambda_1|| (C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t|| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon\}} \\ &\stackrel{a.s.}{\rightarrow} 4 \left(\lambda'_1 J \lambda_1 (C^4 + \mathbb{E}[\eta_t^2]^2) + \lambda_2^2 \right) \times 0 = 0 \end{aligned}$$

To appreciate why the indicator function converges to 0 almost surely we employ (4.27) and (4.76) and note $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ almost surely to get

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n ||\hat{D}_t||^4 &\leq \frac{1}{n} \sum_{t=1}^n \left(||D_t(\hat{\theta}_n)|| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + ||D_t(\hat{\theta}_n)||) \right)^4 \quad (4.77) \\ &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^4 \leq 2^4 \left(\frac{1}{n} \sum_{t=1}^n U_t^4 + \frac{C_1^4}{\underline{\omega}^4} \frac{1}{n} \sum_{t=1}^n \{\rho^t (1 + U_t)\}^4 \right). \end{aligned}$$

The uniform ergodic theorem and Assumption 4.9(ii) imply $\frac{1}{n} \sum_{t=1}^n U_t^4 \xrightarrow{a.s.} \mathbb{E}[U_t^4] < \infty$. Further, (4.28) leads to $\rho^t(1 + U_t) \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$, which in turn implies $\{\rho^t(1 + U_t)\}^4 \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$. Cesàro's lemma yields $\frac{1}{n} \sum_{t=1}^n \{\rho^t(1 + U_t)\}^4 \xrightarrow{a.s.} 0$ and we have $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 < \infty$ almost surely. Thus, $\max_t \|\hat{D}_t\|/\sqrt{n} \xrightarrow{a.s.} 0$ as

$$\left(\frac{\max_t \|\hat{D}_t\|}{\sqrt{n}} \right)^4 \leq \frac{1}{n^2} \sum_{t=1}^n \|\hat{D}_t\|^4 \xrightarrow{a.s.} 0.$$

and $\mathbb{1}_{\{|\lambda_1|(C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t \|\hat{D}_t\| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon\}} \xrightarrow{a.s.} 0$ follows. Combining results, establishes $\frac{1}{s_n^2} \sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}}] \xrightarrow{a.s.} 0$. The Central Limit Theorem for triangular arrays (cf. Billingsley, 1986, Theorem 27.3) implies that $\sum_{t=1}^n Z_{n,t}^*$ converges in conditional distribution to $N(0, \lambda' \Upsilon_\alpha \lambda)$ almost surely. \square

Lemma 4.8. *Suppose Assumptions 4.1–4.9 hold with $a = \pm 6$, $b = 6$ and $c = 2$. Then, we have $I_n^*(z) \xrightarrow{p^*} \frac{z^2}{2} f(\xi_\alpha)$ in probability.*

Proof. Using Fubini's theorem, the conditional expectation is equal to

$$\begin{aligned} \mathbb{E}^*[I_n^*(z)] &= \sum_{t=1}^n \int_0^{z/\sqrt{n}} \mathbb{E}^* \left[\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right] ds \\ &= n \int_0^{z/\sqrt{n}} \left(\hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right) ds \\ &= \int_0^z \sqrt{n} \left(\hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right) du \\ &= \underbrace{\int_0^z \sqrt{n} \left(\hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) - F \left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \right) + F(\hat{\xi}_{n,\alpha}) \right) du}_I \\ &\quad + \underbrace{\int_0^z \sqrt{n} \left(F \left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \right) - F(\hat{\xi}_{n,\alpha}) \right) du}_{II}. \end{aligned}$$

Regarding I , take $\varrho \in (0, 1/2)$ and set $\bar{\mathcal{I}}_n = [\xi_\alpha - 0.5n^{-\varrho}, \xi_\alpha + 0.5n^{-\varrho}]$. Since $\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) = O_p(1)$, the probabilities of the events $\{\hat{\xi}_{n,\alpha} + \frac{|z|}{\sqrt{n}} \notin \bar{\mathcal{I}}_n\}$ and $\{\hat{\xi}_{n,\alpha} - \frac{|z|}{\sqrt{n}} \notin \bar{\mathcal{I}}_n\}$ can be made arbitrarily small for large n . If $\hat{\xi}_{n,\alpha} + \frac{|z|}{\sqrt{n}} \in \bar{\mathcal{I}}_n$ and $\hat{\xi}_{n,\alpha} - \frac{|z|}{\sqrt{n}} \in \bar{\mathcal{I}}_n$, then $\hat{\xi}_{n,\alpha} \in \bar{\mathcal{I}}_n$ and $\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \in \bar{\mathcal{I}}_n$ belong to $\bar{\mathcal{I}}_n$ for all u

between 0 and z . In that case

$$|I| \leq |z| \sup_{x,y \in \tilde{\mathcal{I}}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y)) \right| \xrightarrow{p} 0$$

by Lemma 4.3. Focusing on II , the mean value theorem implies that

$$II = \int_0^z u f(\hat{\xi}_{n,\alpha} + \varepsilon_n) du = \underbrace{\int_0^z u \left(f(\hat{\xi}_{n,\alpha} + \varepsilon_n) - f(\xi_\alpha) \right) du}_{II_1} + \underbrace{\int_0^z u f(\xi_\alpha) du}_{II_2}$$

with ε_n lying between 0 and u/\sqrt{n} . Since $|\varepsilon_n| \leq |z|/\sqrt{n}$ and $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$ we have

$$|II_1| \leq \frac{z^2}{2} \sup_{|v| \leq |z|} \left| f\left(\hat{\xi}_{n,\alpha} + \frac{v}{n}\right) - f(\xi_\alpha) \right| \xrightarrow{a.s.} 0.$$

Further, II_2 simplifies to $II_2 = \frac{z^2}{2} f(\xi_\alpha)$ and combining results establishes

$$\mathbb{E}^*[I_n^*(z)] \xrightarrow{p} \frac{z^2}{2} f(\xi_\alpha).$$

The conditional variance vanishes in probability as

$$\begin{aligned} \text{Var}^*[I_n^*(z)] &= \sum_{t=1}^n \text{Var}^* \left[\int_0^{z/\sqrt{n}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds \right] \\ &= n \text{Var}^* \left[\int_0^{z/\sqrt{n}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds \right] \\ &\leq n \frac{|z|}{\sqrt{n}} \mathbb{E}^* \left[\int_0^{z/\sqrt{n}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds \right] \\ &= \frac{|z|}{\sqrt{n}} \int_0^{z/\sqrt{n}} n \left(\hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha}-) \right) ds \\ &= \frac{|z|}{\sqrt{n}} \mathbb{E}^*[I_n^*(z)] \xrightarrow{p} 0, \end{aligned}$$

where the inequality follows from the fact that

$$\text{Var}(Y) \leq |c| \mathbb{E}[Y] \tag{4.78}$$

with $Y = \int_0^c (\mathbb{1}_{\{X \leq s\}} - \mathbb{1}_{\{X < 0\}}) ds$, X is a real-valued integrable random variable and $c \in \mathbb{R}$ (cf. Francq and Zakoïan, 2015, p. 171). \square

Lemma 4.9. *Suppose Assumptions 4.1–4.10 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have $J_{n,1}^*(z) \xrightarrow{d^*} \Gamma(\frac{r}{2}, \frac{\kappa-1}{4}\xi_\alpha^2 f(\xi_\alpha))$ in probability, i.e. a Gamma distribution with shape parameter $\frac{r}{2}$ and scale parameter $\frac{\kappa-1}{4}\xi_\alpha^2 f(\xi_\alpha)$.*

Proof. We set $\bar{\xi}_{n,\alpha}(z) = \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}$ and define

$$\begin{aligned} T_n^*(z, u) &= \sum_{t=1}^n \tau_{n,t}^*(z, u) \\ \tau_t^*(z, u) &= \int_0^{(1-\tilde{\lambda}_t^{-1}(u))\eta_t^*} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}}) ds \\ \tilde{\lambda}_t(u) &= \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}u)}{\tilde{\sigma}_t(\hat{\theta}_n)} \end{aligned}$$

for $z \in \mathbb{R}$ and $u \in \mathbb{R}^r$, where we suppress the dependence of $\tau_t^*(z, u)$ and $\tilde{\lambda}_t(u)$ on n for notational simplicity. We decompose $T_n^*(z, u) = T_{n,1}^*(z, u) + T_{n,2}^*(z, u)$ with

$$T_{n,1}^*(z, u) = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tau_t^*(z, u) \quad \text{and} \quad T_{n,2}^*(z, u) = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) < 1\}} \tau_t^*(z, u).$$

Let $A > 0$; We establish the result in three steps:

$$\text{Step 1: } T_{n,1}^*(z, u) \xrightarrow{p^*} \frac{1}{2}\xi_\alpha^2 f(\xi_\alpha) \mathbb{E} \left[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u \right]$$

$$T_{n,2}^*(z, u) \xrightarrow{p^*} \frac{1}{2}\xi_\alpha^2 f(\xi_\alpha) \mathbb{E} \left[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u \right]$$

in probability for all $z \in \mathbb{R}$ and for all $u \in \{u \in \mathbb{R}^r : \|u\| \leq A\}$;

$$\text{Step 2: } \sup_{\|u\| \leq A} |T_n^*(z, u) - \frac{1}{2}\xi_\alpha^2 f(\xi_\alpha) u' J u| \xrightarrow{p^*} 0 \text{ in probability for all } z \in \mathbb{R};$$

$$\text{Step 3: } J_{n,1}^*(z) \xrightarrow{d^*} \Gamma(\frac{r}{2}, \frac{\kappa-1}{4}\xi_\alpha^2 f(\xi_\alpha)) \text{ in probability.}$$

Consider *Step 1*; employing the elementary equality

$$\int_0^c (\mathbb{1}_{\{x \leq s\}} - \mathbb{1}_{\{x < 0\}}) ds = (x - c)(\mathbb{1}_{\{c \leq x < 0\}} - \mathbb{1}_{\{0 \leq x < c\}}), \quad (4.79)$$

for $c, s, x \in \mathbb{R}$ and rearranging we obtain

$$\begin{aligned} \tau_t^*(z, u) &= \left((\eta_t^* - \bar{\xi}_{n,\alpha}(z)) - (1 - \tilde{\lambda}_t^{-1}(u))\eta_t^* \right) \\ &\quad \times \left(\mathbb{1}_{\{(1-\tilde{\lambda}_t^{-1}(u))\eta_t^* \leq \eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}} - \mathbb{1}_{\{0 < \eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq (1-\tilde{\lambda}_t^{-1}(u))\eta_t^*\}} \right) \end{aligned}$$

$$\begin{aligned}
 &= \tilde{\lambda}_t^{-1}(u) \left((\eta_t^* - \bar{\xi}_{n,\alpha}(z)) - (1 - \tilde{\lambda}_t(u)) \bar{\xi}_{n,\alpha}(z) \right) \\
 &\quad \times \left(\mathbb{1}_{\{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z) \leq \eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}} - \mathbb{1}_{\{0 < \eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq (1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)\}} \right) \\
 &= \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}} \right) ds. \tag{4.80}
 \end{aligned}$$

Hence, we have

$$T_{n,1}^*(z, u) = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}} \right) ds.$$

Using Fubini's theorem, its conditional mean is equal to

$$\sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) -) \right) ds,$$

which can be decomposed into

$$\mathbb{E}^*[T_{n,1}^*(z, u)] = I + II + III$$

with

$$\begin{aligned}
 I &= \frac{1}{2} \bar{\xi}_{n,\alpha}^2(z) f(\xi_\alpha) \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) (\tilde{\lambda}_t(u) - 1)^2 \\
 II &= \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(F(\bar{\xi}_{n,\alpha}(z) + s) \right. \\
 &\quad \left. - F(\bar{\xi}_{n,\alpha}(z)) - s f(\xi_\alpha) \right) ds \\
 III &= \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) -) \right. \\
 &\quad \left. - F(\bar{\xi}_{n,\alpha}(z) + s) + F(\bar{\xi}_{n,\alpha}(z)) \right) ds.
 \end{aligned}$$

We consider the first term, which expands to

$$I = I_1(I_2 + I_3 + I_4 + I_5),$$

where

$$I_1 = \frac{1}{2} \bar{\xi}_{n,\alpha}^2(z) f(\xi_\alpha)$$

$$\begin{aligned}
 I_2 &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} (\tilde{\lambda}_t^{-1}(u) - 1) n (\tilde{\lambda}_t(u) - 1)^2 \\
 I_3 &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \left(n (\tilde{\lambda}_t(u) - 1)^2 - u' \hat{D}_t \hat{D}_t' u \right) \\
 I_4 &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} u' (\hat{D}_t \hat{D}_t' - D_t D_t') u \\
 I_5 &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} u D_t D_t' u.
 \end{aligned}$$

Theorem 4.1 gives $\bar{\xi}_{n,\alpha}(z) = \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \xrightarrow{a.s.} \xi_\alpha$ and $I_1 \xrightarrow{a.s.} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha)$ follows. Regarding I_2 , we employ the mean value theorem as well as (4.27) and (4.74) and obtain

$$\begin{aligned}
 & \sup_{\|u\| \leq A} \sqrt{n} |\tilde{\lambda}_t(u) - 1| = \sup_{\|u\| \leq A} \sqrt{n} \left| \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}u)}{\tilde{\sigma}_t(\hat{\theta}_n)} - 1 \right| \\
 &= \sup_{\|u\| \leq A} \sqrt{n} \left| \frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \frac{\partial \tilde{\sigma}_t(\bar{\theta}_n)}{\partial \theta'} \frac{u}{\sqrt{n}} \right| = \sup_{\|u\| \leq A} \left| \frac{\tilde{\sigma}_t(\bar{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \tilde{D}_t(\bar{\theta}_n) u \right| \\
 &\leq \sup_{\|u\| \leq A} \|u\| \frac{\tilde{\sigma}_t(\bar{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \|\tilde{D}_t(\bar{\theta}_n)\| \tag{4.81} \\
 &\leq \sup_{\|u\| \leq A} \|u\| \left(\frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\hat{\theta}_n)} + \frac{C_1 \rho^t}{\underline{\omega}} \left(1 + \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\hat{\theta}_n)} \right) \right) \\
 &\quad \times \left(\|D_t(\bar{\theta}_n)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\bar{\theta}_n)\|) \right) \\
 &\stackrel{a.s.}{\leq} A \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)
 \end{aligned}$$

with $\bar{\theta}_n$ lying between $\hat{\theta}_n$ and $\hat{\theta}_n + n^{-1/2}u$. Similarly, we find

$$\sup_{\|u\| \leq A} \sqrt{n} |\tilde{\lambda}_t^{-1}(u) - 1| \tag{4.82}$$

$$\stackrel{a.s.}{\leq} A \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right). \tag{4.83}$$

Together with (4.76), we thus have

$$\sqrt{n} |I_2| \leq \frac{1}{n} \sum_{t=1}^n \sqrt{n} |\tilde{\lambda}_t^{-1}(u) - 1| n (\tilde{\lambda}_t(u) - 1)^2$$

$$\begin{aligned}
 &\stackrel{a.s.}{\leq} \frac{A^3}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^3 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^3 \quad (4.84) \\
 &\leq \frac{2^3 A^3}{n} \sum_{t=1}^n \left(S_t^3 T_t^3 + \frac{C_1^3 \rho^{3t}}{\underline{\omega}^3} (1 + S_t T_t)^3 \right) \left(U_t^3 + \frac{C_1^3 \rho^{3t}}{\underline{\omega}^3} (1 + U_t)^3 \right) \\
 &\leq 2^3 A^3 \left(\frac{1}{n} \sum_{t=1}^n S_t^3 T_t^3 U_t^3 + \left(\frac{2C_1^3}{\underline{\omega}^3} + \frac{C_1^6}{\underline{\omega}^6} \right) \frac{1}{n} \sum_{t=1}^n \rho^{3t} (1 + S_t T_t)^3 (1 + U_t)^3 \right).
 \end{aligned}$$

The right-hand side quantity is $O(1)$ almost surely. To appreciate why note that Hölder's inequality, the uniform ergodic theorem and Assumption 4.9 imply

$$\frac{1}{n} \sum_{t=1}^n S_t^3 T_t^3 U_t^3 \leq \underbrace{\left(\frac{1}{n} \sum_{t=1}^n S_t^{12} \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[S_t^{12}] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^{12} \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^{12}] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^6 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^6] < \infty}.$$

Further, for each $\varepsilon > 0$, Markov's inequality yields

$$\begin{aligned}
 &\sum_{t=1}^{\infty} \mathbb{P} \left[\rho^{3t} (1 + S_t T_t)^3 (1 + U_t)^3 > \varepsilon \right] \\
 &\leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t T_t)(1 + U_t)]}{\varepsilon^{1/3}} = \frac{\mathbb{E}[(1 + S_t T_t)(1 + U_t)]}{\varepsilon^{1/3} (1 - \rho)} < \infty
 \end{aligned}$$

and $\frac{1}{n} \sum_{t=1}^n \rho^{3t} (1 + S_t T_t)^3 (1 + U_t)^3 \xrightarrow{a.s.} 0$ follows from combining the Borel-Cantelli lemma with Cesàro's lemma. Thus, we have $\sqrt{n} I_2 = O(1)$ almost surely implying $I_2 \xrightarrow{a.s.} 0$. Regarding I_3 we use the mean value theorem as well as (4.32) and (4.74)

to find

$$\begin{aligned}
 & \sup_{\|u\| \leq A} \left| \sqrt{n}(\tilde{\lambda}_t(u) - 1) - \hat{D}'_t u \right| \\
 &= \sup_{\|u\| \leq A} \frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \left| \sqrt{n} \left(\tilde{\sigma}_t(\hat{\theta}_n + u/\sqrt{n}) - \tilde{\sigma}_t(\hat{\theta}_n) \right) - \frac{\partial \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta'} u \right| \\
 &= \sup_{\|u\| \leq A} \frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \frac{1}{\sqrt{n}} u' \frac{\partial^2 \tilde{\sigma}_t(\check{\theta}_n)}{\partial \theta \partial \theta'} u \leq \sup_{\|u\| \leq A} \|u\|^2 \frac{1}{\sqrt{n}} \frac{\tilde{\sigma}_t(\check{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \|\tilde{H}(\check{\theta}_n)\| \\
 &\leq \sup_{\|u\| \leq A} \|u\|^2 \frac{1}{\sqrt{n}} \left(\frac{\sigma_t(\check{\theta}_n)}{\sigma_t(\hat{\theta}_n)} + \frac{C_1 \rho^t}{\underline{\omega}} \left(1 + \frac{\sigma_t(\check{\theta}_n)}{\sigma_t(\hat{\theta}_n)} \right) \right) \\
 &\quad \times \left(\|H(\check{\theta}_n)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|H(\check{\theta}_n)\|) \right) \\
 &\stackrel{a.s.}{\leq} \frac{A^2}{\sqrt{n}} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right)
 \end{aligned} \tag{4.85}$$

with $\check{\theta}_n$ lying between $\hat{\theta}_n$ and $\hat{\theta}_n + n^{-1/2}u$. Together with (4.76), we establish

$$\begin{aligned}
 |I_3| &\leq \frac{1}{n} \sum_{t=1}^n \left| n(\tilde{\lambda}_t(u) - 1)^2 - (\hat{D}'_t u)^2 \right| \\
 &\leq \frac{1}{n} \sum_{t=1}^n \left(\sqrt{n}(\tilde{\lambda}_t(u) - 1) - u' \hat{D}_t \right)^2 + \frac{2}{n} \sum_{t=1}^n |\hat{D}'_t u| \left| \sqrt{n}(\tilde{\lambda}_t(u) - 1) - u' \hat{D}_t \right| \\
 &\stackrel{a.s.}{\leq} \frac{A^4}{n^2} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^2 \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right)^2 \\
 &\quad + \frac{2A^3}{n\sqrt{n}} \sum_{t=1}^n U_t \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right) \\
 &\leq \frac{A^4}{n^2} \sum_{t=1}^n \left(S_t^2 T_t^2 + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + S_t T_t)^2 \right) \left(V_t^2 + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + V_t)^2 \right) \\
 &\quad + \frac{2A^3}{n\sqrt{n}} \sum_{t=1}^n U_t \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right) \\
 &\leq \frac{A^4}{n} \left(\frac{1}{n} \sum_{t=1}^n S_t^2 T_t^2 V_t^2 + \left(\frac{2C_1^2}{\underline{\omega}^2} + \frac{C_1^4}{\underline{\omega}^4} \right) \frac{1}{n} \sum_{t=1}^n \rho^{2t} (1 + S_t T_t)^2 (1 + V_t)^2 \right) \\
 &\quad + \frac{2A^3}{\sqrt{n}} \left(\frac{1}{n} \sum_{t=1}^n S_t T_t U_t V_t + \left(\frac{2C_1}{\underline{\omega}} + \frac{C_1^2}{\underline{\omega}^2} \right) \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t T_t) U_t (1 + V_t) \right)
 \end{aligned}$$

Hölder's inequality, the uniform ergodic theorem and Assumption 4.9 imply

$$\frac{1}{n} \sum_{t=1}^n S_t^2 T_t^2 V_t^2 \leq \underbrace{\left(\frac{1}{n} \sum_{t=1}^n S_t^8 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[S_t^8] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^8 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^8] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n V_t^4 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[V_t^4] < \infty}$$

and

$$\frac{1}{n} \sum_{t=1}^n S_t T_t U_t V_t \leq \underbrace{\left(\frac{1}{n} \sum_{t=1}^n S_t^4 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[S_t^4] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n T_t^4 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^4] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^4 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^4] < \infty} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n V_t^4 \right)^{\frac{1}{4}}}_{\xrightarrow{a.s.} \mathbb{E}[V_t^4] < \infty}.$$

Further, for each $\varepsilon > 0$, Markov's inequality yields

$$\begin{aligned} & \sum_{t=1}^{\infty} \mathbb{P} \left[\rho^{2t} (1 + S_t T_t)^2 (1 + V_t)^2 > \varepsilon \right] \\ & \leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t T_t)(1 + V_t)]}{\sqrt{\varepsilon}} = \frac{\mathbb{E}[(1 + S_t T_t)(1 + V_t)]}{\sqrt{\varepsilon}(1 - \rho)} < \infty \end{aligned}$$

implying $\frac{1}{n} \sum_{t=1}^n \rho^{2t} (1 + S_t T_t)^2 (1 + V_t)^2 \xrightarrow{a.s.} 0$ and

$$\begin{aligned} & \sum_{t=1}^{\infty} \mathbb{P} \left[\rho^t (1 + S_t T_t) U_t (1 + V_t) > \varepsilon \right] \\ & \leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t T_t) U_t (1 + V_t)]}{\varepsilon} = \frac{\mathbb{E}[(1 + S_t T_t) U_t (1 + V_t)]}{\varepsilon(1 - \rho)} < \infty. \end{aligned}$$

yielding $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t T_t) U_t (1 + V_t) \xrightarrow{a.s.} 0$. Combining results, we have $I_3 \xrightarrow{a.s.} 0$.

With regard to I_4 the mean value theorem implies

$$\begin{aligned} & \left\| D_t(\hat{\theta}_n) - D_t(\theta_0) \right\| = \left\| \frac{1}{\sigma_t(\hat{\theta}_n)} \frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta} - \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \right\| \\ & = \left\| \left(\frac{1}{\sigma_t(\hat{\theta}_n)} \frac{\partial^2 \sigma_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} - \frac{1}{\sigma_t^2(\hat{\theta}_n)} \frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta'} \right) (\hat{\theta}_n - \theta_0) \right\| \quad (4.86) \\ & = \left\| \left(H_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n) \right) (\hat{\theta}_n - \theta_0) \right\| \stackrel{a.s.}{\leq} (V_t + U_t^2) \|\hat{\theta}_n - \theta_0\| \end{aligned}$$

with $\dot{\theta}_n$ between $\hat{\theta}_n$ and θ_0 . The triangle inequality and (4.31) lead to

$$\begin{aligned}
 |I_4| &\leq \|u\|^2 \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t \hat{D}'_t - D_t D'_t\| \\
 &\leq \frac{A^2}{n} \sum_{t=1}^n \left\| D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) - D_t(\theta_0) D'_t(\theta_0) \right\| \\
 &\quad + \frac{A^2}{n} \sum_{t=1}^n \left\| \tilde{D}_t(\hat{\theta}_n) \tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) \right\| \\
 &\leq \frac{A^2}{n} \sum_{t=1}^n \|D_t(\hat{\theta}_n) - D_t\|^2 + \frac{2A^2}{n} \sum_{t=1}^n \|D_t\| \|D_t(\hat{\theta}_n) - D_t\| \\
 &\quad + \frac{A^2}{n} \sum_{t=1}^n \left\| \tilde{D}_t(\hat{\theta}_n) \tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) \right\| \\
 &\stackrel{a.s.}{\leq} A^2 \|\hat{\theta}_n - \theta_0\|^2 \frac{1}{n} \sum_{t=1}^n (V_t + U_t^2)^2 + 2A^2 \|\hat{\theta}_n - \theta_0\| \frac{1}{n} \sum_{t=1}^n U_t (V_t + U_t^2) \\
 &\quad + A^2 \left(\frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t)^2.
 \end{aligned}$$

Employing Hölder's inequality, the uniform ergodic theorem and Assumption 4.9 one can show that $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (V_t + U_t^2)^2 = O(1)$ almost surely as well as $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n U_t (V_t + U_t^2) = O(1)$ almost surely. Together with $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ the first and the second term vanish. The third term vanishes almost surely by the discussion below (4.31), which establishes $I_4 \xrightarrow{a.s.} 0$. Next, we show that I_5 converges in probability to $\mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$. For $\varepsilon > 0$ we have

$$\begin{aligned}
 I_5 &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\sqrt{n}(\bar{\lambda}_t(u) - 1) > -\varepsilon\}} u' D_t D'_t u \\
 &\leq \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{D'_t u > -2\varepsilon\}} u' D_t D'_t u}_{\xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{\{D'_t u > -2\varepsilon\}} u' D_t D'_t u]} + \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\bar{\lambda}_t(u) - 1) - D'_t u| > \varepsilon\}} u' D_t D'_t u
 \end{aligned}$$

by the ergodic theorem and similarly

$$I_5 \geq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\sqrt{n}(\bar{\lambda}_t(u) - 1) > \varepsilon\}} u' D_t D'_t u$$

$$\geq \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{D'_t u > 2\varepsilon\}} u' D_t D'_t u - \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} u' D_t D'_t u}_{\xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{\{D'_t u > 2\varepsilon\}} u' D_t D'_t u]}$$

We can choose $\varepsilon > 0$ sufficiently small such that $\mathbb{E}[\mathbb{1}_{\{D'_t u > \pm 2\varepsilon\}} u' D_t D'_t u]$ is sufficiently close to $\mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$. Given $\varepsilon > 0$, it remains to show that the term $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} u' D_t D'_t u$ vanishes in probability. The Cauchy-Schwarz inequality, the ergodic theorem and Assumption 4.9 imply

$$\begin{aligned} 0 &\leq \sup_{\|u\| \leq A} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} u' D_t D'_t u \\ &\leq A^2 \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} U_t^2 \\ &\leq A^2 \left(\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} \right)^{\frac{1}{2}} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n U_t^4 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^4] < \infty} \end{aligned}$$

We are left to show that the average in the first brackets vanishes in probability. Since the term is positive, it suffices to prove that its expected value vanishes as n grows large. For every $C > 0$, we obtain

$$\begin{aligned} &\mathbb{E} \left[\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon\}} \right] \\ &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[|\sqrt{n}(\tilde{\lambda}_t(u)-1)-D'_t u| > \varepsilon \cap \sqrt{n} \|\hat{\theta}_n - \theta_0\| \leq C \right] \\ &\quad + \mathbb{P} \left[\sqrt{n} \|\hat{\theta}_n - \theta_0\| > C \right]. \end{aligned}$$

The second term can be made arbitrarily small by choosing C sufficiently large.

Given C , we use (4.27), (4.85) and (4.86) as well as Markov's inequality to obtain

$$\begin{aligned}
 & \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[\left| \sqrt{n}(\tilde{\lambda}_t(u) - 1) - D'_t u \right| > \varepsilon \cap \|\hat{\theta}_n - \theta_0\| \leq C/\sqrt{n} \right] \\
 & \leq \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[\frac{A^2}{\sqrt{n}} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right) \right. \\
 & \quad \left. + A \left(\frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) + (V_t + U_t^2) \frac{C}{\sqrt{n}} \right) > \varepsilon \right] \\
 & \leq \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[\frac{A}{n^{1/4}} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) > \frac{\sqrt{\varepsilon}}{2} \right] \\
 & \quad + \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[A \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) > \frac{\varepsilon}{4} \right] \\
 & \quad + \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[\frac{A}{n^{1/4}} \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right) > \frac{\sqrt{\varepsilon}}{2} \right] \\
 & \quad + \frac{1}{n} \sum_{t=1}^n \mathbb{P} \left[\frac{AC}{\sqrt{n}} (V_t + U_t^2) > \frac{\varepsilon}{4} \right] \\
 & \leq \left(\frac{2A}{\sqrt{\varepsilon} n^{1/4}} \right)^{\frac{s}{2}} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[\left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^{\frac{s}{2}} \right] \\
 & \quad + \left(\frac{4A}{\varepsilon \underline{\omega}} \right)^{\frac{s}{2}} \frac{1}{n} \sum_{t=1}^n \rho^{\frac{st}{2}} \mathbb{E} \left[C_1^{\frac{s}{2}} (1 + U_t)^{\frac{s}{2}} \right] \\
 & \quad + \left(\frac{2A}{\sqrt{\varepsilon} n^{1/4}} \right)^{\frac{s}{2}} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[\left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) \right)^{\frac{s}{2}} \right] \\
 & \quad + \frac{4AC}{\varepsilon \sqrt{n}} \frac{1}{n} \sum_{t=1}^n \mathbb{E} [V_t + U_t^2].
 \end{aligned} \tag{4.87}$$

We have $\frac{1}{n} \sum_{t=1}^n \mathbb{E} [V_t + U_t^2] = \mathbb{E} [V_t] + \mathbb{E} [U_t^2]$ and using Hölder's inequality as well as Assumptions 4.4 and 4.9, it is straight-forward to show that $\frac{1}{n} \sum_{t=1}^n \rho^{\frac{st}{2}} \mathbb{E} [C_1^{\frac{s}{2}} (1 + U_t)^{\frac{s}{2}}] \rightarrow 0$, $\frac{1}{n} \sum_{t=1}^n \mathbb{E} [(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t))^{\frac{s}{2}}] < \infty$ and $\frac{1}{n} \sum_{t=1}^n \mathbb{E} [(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t))^{\frac{s}{2}}] < \infty$ such that (4.87) goes to 0. Therefore, we obtain that the term $\sup_{\|u\| \leq A} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\sqrt{n}(\tilde{\lambda}_t(u) - 1) - D'_t u| > \varepsilon\}} u' D_t D'_t u$ converges in probability to 0 completing the proof of $I_5 \xrightarrow{P} \mathbb{E} [\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$. Having analyzed I_1, \dots, I_5 we conclude that $I \xrightarrow{P} \frac{1}{2} \xi_\alpha f(\xi_\alpha) \mathbb{E} [\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$.

Consider II ; the mean value theorem yields

$$II = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} s \left(f(\bar{\xi}_{n,\alpha}(z) + \varepsilon_{t,n}(z, u)) - f(\xi_\alpha) \right) ds$$

with $\varepsilon_{t,n}(z, u)$ between 0 and $(1 - \tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)$. In addition, we have

$$n^{1/8} \max_{t=1, \dots, n} \sup_{\|u\| \leq A} |\tilde{\lambda}_t(u) - 1| \xrightarrow{P} 0. \quad (4.88)$$

To appreciate why, note that (4.81) gives

$$\begin{aligned} & n^{1/8} \max_{t=1, \dots, n} \sup_{\|u\| \leq A} |\tilde{\lambda}_t(u) - 1| \\ & \leq \frac{a.s.}{n^{3/8}} \max_{t=1, \dots, n} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right). \end{aligned}$$

The right-hand side converges in probability to zero since for every $\varepsilon > 0$

$$\begin{aligned} & \mathbb{P} \left[\frac{A}{n^{3/8}} \max_{t=1, \dots, n} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right) \geq \varepsilon \right] \\ & = \mathbb{P} \left[\frac{A^3}{n^{9/8}} \max_{t=1, \dots, n} \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^3 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^3 \geq \varepsilon^3 \right] \\ & \leq \mathbb{P} \left[\frac{A^3}{n^{9/8}} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^3 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^3 \geq \varepsilon^3 \right] \rightarrow 0 \end{aligned}$$

with regard to (4.84). We define

$$\begin{aligned} \bar{\xi}_{n,\alpha}^+(z) &= \bar{\xi}_{n,\alpha}(z) + \max_{t=1, \dots, n} |\tilde{\lambda}_t(u) - 1| |\bar{\xi}_{n,\alpha}(z)| \\ \bar{\xi}_{n,\alpha}^-(z) &= \bar{\xi}_{n,\alpha}(z) - \max_{t=1, \dots, n} |\tilde{\lambda}_t(u) - 1| |\bar{\xi}_{n,\alpha}(z)|. \end{aligned}$$

and set $\mathcal{I}_n = [\xi_\alpha - a_n, \xi_\alpha + a_n]$ with $a_n \sim n^{-1/8} \log n$ as $n \rightarrow \infty$. Since $n^{1/8}(\bar{\xi}_{n,\alpha}^+(z) - \xi_\alpha) \xrightarrow{P} 0$ and $n^{1/8}(\bar{\xi}_{n,\alpha}^-(z) - \xi_\alpha) \xrightarrow{P} 0$ as $\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) = O_p(1)$ and (4.88) holds, the probabilities of the events $\{\bar{\xi}_{n,\alpha}^+(z) \notin \mathcal{I}_n\}$ and $\{\bar{\xi}_{n,\alpha}^-(z) \notin \mathcal{I}_n\}$ can be made arbitrarily small for large n . If $\hat{\xi}_{n,\alpha}^+, \hat{\xi}_{n,\alpha}^- \in \mathcal{I}_n$, then $\bar{\xi}_{n,\alpha}(z) + \varepsilon_{t,n}(z, u)$ belong to \mathcal{I}_n for all $t = 1, \dots, n$, for all u such that $\|u\| \leq A$ and given z . In that

case we have

$$|II| \leq \frac{1}{2} \bar{\xi}_{n,\alpha}^2(z) \sup_{x \in \mathcal{I}_n} |f(x) - f(\xi_\alpha)| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) n(1 - \tilde{\lambda}_t(u))^2.$$

We have $\bar{\xi}_{n,\alpha}^2(z) \xrightarrow{a.s.} \xi_\alpha^2$ by Theorem 4.1 and $\sup_{x \in \mathcal{I}_n} |f(x) - f(\xi_\alpha)| \rightarrow 0$ as \mathcal{I}_n shrinks to ξ_α and continuity of f in a neighborhood of ξ_α (see Assumption 4.4(ii)). Moreover, $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-1}(u) n(1 - \tilde{\lambda}_t(u))^2 = I_2 + I_3 + I_4 + I_5 \xrightarrow{P} \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$ and we conclude $|II| \xrightarrow{P} 0$.

Focusing on III , we only consider the case $\hat{\xi}_{n,\alpha}^+, \hat{\xi}_{n,\alpha}^- \in \mathcal{I}_n$ since the probabilities of the events $\{\bar{\xi}_{n,\alpha}^+(z) \notin \mathcal{I}_n\}$ and $\{\bar{\xi}_{n,\alpha}^-(z) \notin \mathcal{I}_n\}$ can be made arbitrarily small for large n . Because $\hat{\xi}_{n,\alpha}^+, \hat{\xi}_{n,\alpha}^- \in \mathcal{I}_n$ implies that $\bar{\xi}_{n,\alpha}(z)$ and $\bar{\xi}_{n,\alpha}(z) + s$ belong to \mathcal{I}_n for all s between 0 and $(1 - \tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)$ for all $t = 1, \dots, n$, we obtain

$$\begin{aligned} |III| &\leq |\bar{\xi}_{n,\alpha}(z)| \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y)) \right| \\ &\quad \times \frac{1}{n} \sum_{t=1}^n \sqrt{n} |\tilde{\lambda}_t^{-1}(u) - 1|. \end{aligned}$$

We have $|\bar{\xi}_{n,\alpha}(z)| \xrightarrow{a.s.} |\xi_\alpha|$ and $\sup_{x,y \in \mathcal{I}_n} |\sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y))| \xrightarrow{P} 0$ by Lemma 4.3. Moreover, (4.82) and (4.84) together with Hölder's inequality yield

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \sqrt{n} |\tilde{\lambda}_t^{-1}(u) - 1| \\ &\leq \frac{1}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right) \quad (4.89) \\ &\leq \left(\frac{1}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^3 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^3 \right)^{\frac{1}{3}} = O(1) \end{aligned}$$

almost surely as $n \rightarrow \infty$. We conclude $III \xrightarrow{P} 0$ and establish

$$\mathbb{E}[T_{n,1}^*(z, u)] \xrightarrow{P} \frac{1}{2} \xi_\alpha f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]. \quad (4.90)$$

Employing (4.78), the conditional $T_{n,1}^*(z, u)$ is bounded by

$$\mathbb{V}\text{ar}^*[T_{n,1}^*(z, u)]$$

$$\begin{aligned}
 &= \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t(u) > 1\}} \tilde{\lambda}_t^{-2}(u) \\
 &\quad \times \mathbb{V}\text{ar}^* \left[\int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}}) ds \right] \\
 &\leq \sum_{t=1}^n \tilde{\lambda}_t^{-2}(u) |\tilde{\lambda}_t(u) - 1| |\bar{\xi}_{n,\alpha}(z)| \\
 &\quad \times \mathbb{E}^* \left[\int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha}(z) < 0\}}) ds \right] \\
 &= |\bar{\xi}_{n,\alpha}(z)| \sum_{t=1}^n \tilde{\lambda}_t^{-2}(u) |\tilde{\lambda}_t(u) - 1| \\
 &\quad \times \int_0^{(1-\tilde{\lambda}_t(u))\bar{\xi}_{n,\alpha}(z)} \left(\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}(z) -) \right) ds \\
 &\leq \bar{\xi}_{n,\alpha}^2(z) \sum_{t=1}^n \tilde{\lambda}_t^{-2}(u) |\tilde{\lambda}_t(u) - 1|^2 \left(\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^+(z)) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^-(z)) \right) \\
 &= \bar{\xi}_{n,\alpha}^2(z) \frac{1}{n} \sum_{t=1}^n n |\tilde{\lambda}_t^{-1}(u) - 1|^2 \left(\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^+(z)) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^-(z)) \right).
 \end{aligned}$$

We have $\bar{\xi}_{n,\alpha}^2(z) \xrightarrow{a.s.} \xi_\alpha^2$ and $\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^+(z)) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^-(z)) \xrightarrow{P} F(\xi_\alpha) - F(\xi_\alpha) = 0$ since $\bar{\xi}_{n,\alpha}^+(z) \xrightarrow{P} \xi_\alpha$, $\bar{\xi}_{n,\alpha}^-(z) \xrightarrow{P} \xi_\alpha$ and $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$ (Lemma 4.1). Analogously to (4.89), we find

$$\begin{aligned}
 &\frac{1}{n} \sum_{t=1}^n n |\tilde{\lambda}_t^{-1}(u) - 1|^2 \\
 &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^2 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^2 \\
 &\leq \left(\frac{1}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right)^3 \left(U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^3 \right)^{\frac{2}{3}} = O(1)
 \end{aligned}$$

almost surely as $n \rightarrow \infty$. We conclude $\mathbb{V}\text{ar}^*[T_{n,1}^*(z, u)] \xrightarrow{P} 0$ and together with (4.90) we establish $T_{n,1}^*(z, u) \xrightarrow{P^*} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$. The proof of $T_{n,2}^*(z, u) \xrightarrow{P^*} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u]$ in probability is analogous and hence omitted.

Regarding *Step 2* the triangle inequality yields

$$\begin{aligned}
 & \sup_{||u|| \leq A} \left| T_n^*(z, u) - \text{p} \lim_{n \rightarrow \infty} T_n^*(z, u) \right| \\
 & \leq \sup_{||u|| \leq A} \left| T_{n,1}^*(z, u) - \text{p} \lim_{n \rightarrow \infty} T_{n,1}^*(z, u) \right| \\
 & \quad + \sup_{||u|| \leq A} \left| T_{n,2}^*(z, u) - \text{p} \lim_{n \rightarrow \infty} T_{n,2}^*(z, u) \right|.
 \end{aligned} \tag{4.91}$$

Let $N \geq 1$ be an integer. Analogously to the third step in Lemma 4.3, we divide the (hyper-)cube $[-A, A]^r$ into $L = (2N)^r$ cubes with side length A/N . Again, in case of a cube ℓ , $u_\bullet(\ell)$ and $u^\bullet(\ell)$ denote the lower left and upper right vertex of ℓ . Given $\ell \in \{1, \dots, L\}$ and u satisfying $u_\bullet(\ell) \leq u \leq u^\bullet(\ell)$ (element-by-element comparison), Assumption 4.8 implies $\tilde{\lambda}_t(u_\bullet(\ell)) \leq \tilde{\lambda}_t(u) \leq \tilde{\lambda}_t(u^\bullet(\ell))$. Further, given z , Theorem 4.1 results in $\tilde{\xi}_{n,\alpha}(z) \xrightarrow{a.s.} \xi_\alpha < 0$. Thus, we have for n sufficiently large

$$\begin{aligned}
 T_{n,1}^*(z, u_\bullet(\ell)) & \leq T_{n,1}^*(z, u) \leq T_{n,1}^*(z, u^\bullet(\ell)) \\
 T_{n,2}^*(z, u^\bullet(\ell)) & \leq T_{n,2}^*(z, u) \leq T_{n,2}^*(z, u_\bullet(\ell)).
 \end{aligned}$$

Let $k \in \{1, 2\}$; we obtain

$$\begin{aligned}
 & \sup_{||u|| \leq A} \left| T_{n,k}^*(z, u) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u) \right| \\
 & \leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right| \\
 & \quad + \underbrace{\max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \left| T_{n,k}^*(z, u^\bullet(\ell)) - T_{n,k}^*(z, u) \right|}_{A_n} \\
 & \quad + \underbrace{\max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \left| \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u) \right|}_{B_n}
 \end{aligned}$$

with

$$\begin{aligned}
 A_n & \leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - T_{n,k}^*(z, u_\bullet(\ell)) \right| \\
 & \leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right| \\
 & \quad + \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u_\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|
 \end{aligned}$$

$$+ \max_{1 \leq \ell \leq L} \left| \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|$$

$$B_n \leq \max_{1 \leq \ell \leq L} \left| \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|.$$

Hence, we establish the following bound

$$\sup_{\|u\| \leq A} \left| T_{n,k}^*(z, u) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u) \right| \leq 2IV + V + VI$$

with

$$\begin{aligned} IV &= \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right| \\ V &= \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u_\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right| \\ VI &= \max_{1 \leq \ell \leq L} \left| \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|. \end{aligned}$$

Regarding VI , we have for every u satisfying $\|u\| \leq A$ that

$$\text{p} \lim_{n \rightarrow \infty} T_{n,k}^*(z, u) = \begin{cases} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u] & \text{if } k = 1 \\ \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u] & \text{if } k = 2 \end{cases}$$

is continuous in u . Together with $\|u^\bullet(\ell) - u_\bullet(\ell)\| \leq \frac{A}{N}$ for every $\ell \in \{1, \dots, L\}$, it follows that VI can be made arbitrarily small by choosing N sufficiently large. Given N (and L), $IV \xrightarrow{P^*} 0$ in probability and $V \xrightarrow{P^*} 0$ in probability by *Step 1*, which completes *Step 2*.

Consider *Step 3*; for each $\varepsilon > 0$ we obtain

$$\begin{aligned} &\mathbb{P}^* \left[\left| J_{n,1}^*(z) - \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' J \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \right| \geq \varepsilon \right] \\ &\leq \mathbb{P}^* \left[\sup_{\|u\| \leq A} \left| T_n^*(u) - \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) u' J u \right| \geq \varepsilon \right] + \mathbb{P}^* \left[\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > A \right]. \end{aligned}$$

With regard to Proposition 4.1, the second term can be made arbitrarily small for large n by choosing A sufficiently large. Given A , the first term vanishes in probability by *Step 2*. Expanding $\frac{1}{2} = \frac{\kappa-1}{8} \frac{4}{\kappa-1}$, we establish

$$J_{n,1}^*(z) = \frac{\kappa-1}{8} \xi_\alpha^2 f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{4}{\kappa-1} J \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{P^*}(1)$$

in probability. Proposition 4.1 implies that $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)' \frac{4}{\kappa-1} J \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} \chi_r^2$, where χ_r^2 denotes the Chi Square distribution with r degrees of freedom. Further, note that $Y = cQ$ with $c > 0$ and $Q \sim \chi_r^2$ implies $Y \sim \Gamma(r/2, 2c)$. It follows that

$$J_{n,1}^*(z) \xrightarrow{d^*} \Gamma\left(\frac{r}{2}, \frac{\kappa-1}{4} \xi_\alpha^2 f(\xi_\alpha)\right)$$

in probability, which completes *Step 3* and establishes the lemma's claim. \square

Remark 4.5. In the preceding proof of Lemma 4.9 a compactness/supremum argument is employed, in which the monotonicity condition of Assumption 4.8 plays a central role. In contrast, the proof of Francq and Zakoïan (2015, p. 172) rests on a conditional argument involving the density of η_t given $\{\hat{\theta}_n - \theta_0, \eta_u : u < t\}$. This argument does not carry over to the residual bootstrap since the probability mass function of η_t^* given $\{\hat{\theta}_n^* - \hat{\theta}_n, \eta_u^* : u < t\}$ and \mathcal{F}_n has, almost surely, a single point mass.

Lemma 4.10. *Suppose Assumptions 4.1–4.10 with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have $J_{n,2}^*(z) = z \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$ in probability.*

Proof. Inserting $\hat{\eta}_t^* = \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} \eta_t^*$ into (4.21) leads to

$$J_{n,2}^*(z) = \sum_{t=1}^n \left(1 - \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)}\right) \underbrace{\eta_t^* (\mathbb{1}_{\{\eta_t^* < \xi_{n,\alpha} + \frac{z}{\sqrt{n}}\}} - \mathbb{1}_{\{\eta_t^* < \xi_{n,\alpha}\}})}_{j_{n,t}^{*(2)}(z)}. \quad (4.92)$$

A Taylor expansion around $\hat{\theta}_n$ yields

$$\begin{aligned} & 1 - \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} \\ &= \frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \frac{\partial \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta} (\hat{\theta}_n^* - \hat{\theta}_n) \\ & \quad + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \left(\frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \frac{\partial^2 \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} - \frac{2}{\tilde{\sigma}_t^2(\hat{\theta}_n)} \frac{\partial \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta'} \right) (\hat{\theta}_n^* - \hat{\theta}_n) \\ &= \hat{D}_t'(\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \left(\tilde{H}_t(\hat{\theta}_n) - 2 \tilde{D}_t(\hat{\theta}_n) \tilde{D}_t'(\hat{\theta}_n) \right) (\hat{\theta}_n^* - \hat{\theta}_n), \end{aligned} \quad (4.93)$$

where $\check{\theta}_n$ lies between $\hat{\theta}_n^*$ and $\hat{\theta}_n$. Plugging this result into (4.92) gives

$$\begin{aligned} J_{n,2}^*(z) &= \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n j_{n,t}^{*(2)}(z) \hat{D}_t'}_I \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ &\quad + \underbrace{\frac{1}{2} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)' \frac{1}{n} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left(\tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n)\tilde{D}_t'(\check{\theta}_n) \right) j_{n,t}^{*(2)}(z)}_{II} \\ &\quad \times \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n). \end{aligned}$$

With regard to Proposition 4.1, it suffices to show that $I \xrightarrow{p^*} \xi_\alpha z f(\xi_\alpha) \Omega'$ in probability and $II \xrightarrow{p^*} 0$ in probability. The conditional mean and variance of the first term are

$$\begin{aligned} \mathbb{E}^*[I] &= \sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}] \frac{1}{n} \sum_{t=1}^n \hat{D}_t' = \sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}(z)] \hat{\Omega}_n' \\ \text{Var}^*[I] &= \text{Var}^*[j_{n,t}^{*(2)}] \frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t' = \text{Var}^*[j_{n,t}^{*(2)}(z)] \hat{J}_n. \end{aligned} \tag{4.94}$$

Lemma 4.2 states $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$ and $\hat{J}_n \xrightarrow{a.s.} J$. Further, we have $\sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}(z)] \xrightarrow{p} z \xi_\alpha f(\xi_\alpha)$ and $\sqrt{n} \mathbb{E}^*[(j_{n,t}^{*(2)}(z))^2] \xrightarrow{p} |z| \xi_\alpha^2 f(\xi_\alpha)$, which implies $\text{Var}^*[j_{n,t}^{*(2)}(z)] \xrightarrow{p} 0$. To appreciate why, we obtain for $z \geq 0$

$$\begin{aligned} \sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}(z)] &= \sqrt{n} \int_{[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}})} x d\hat{\mathbb{F}}_n(x) \\ &= \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right) \sqrt{n} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\xi}_{n,\alpha} \sqrt{n} \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \\ &\quad - \sqrt{n} \int_{[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}})} \hat{\mathbb{F}}_n(x) dx \\ &= \underbrace{\hat{\xi}_{n,\alpha} \sqrt{n} \left(\hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right)}_{I_1} + \underbrace{z \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right)}_{I_2} \\ &\quad - \underbrace{\int_{[0,z)} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy}_{I_3}. \end{aligned}$$

Using Lemma 4.3 and the mean value theorem, we find

$$\begin{aligned} I_1 &= \hat{\xi}_{n,\alpha} \sqrt{n} \left(F \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - F(\hat{\xi}_{n,\alpha}) \right) + o_p(1) \\ &= z \hat{\xi}_{n,\alpha} f(\hat{\xi}_{n,\alpha} + \varepsilon_n) + o_p(1), \end{aligned}$$

where $0 \leq \varepsilon_n \leq z/\sqrt{n}$, and together with Theorem 4.1 we establish $I_1 \xrightarrow{P} z \xi_\alpha f(\xi_\alpha)$. Moreover, Theorem 4.1 and Lemma 4.1 imply $I_2 \xrightarrow{P} z F(\xi_\alpha)$ and using additionally the dominated convergence theorem, we obtain $I_3 \xrightarrow{P} z F(\xi_\alpha)$. Hence, $\sqrt{n} \mathbb{E}^* [j_{n,t}^{*(2)}(z)] \xrightarrow{P} z \xi_\alpha f(\xi_\alpha)$ for $z \geq 0$ and analogously one can show it to hold for $z < 0$. Similarly, we find for $z \geq 0$

$$\begin{aligned} \sqrt{n} \mathbb{E}^* \left[(j_{n,t}^{*(2)}(z))^2 \right] &= \sqrt{n} \int_{[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}})} x^2 d\hat{\mathbb{F}}_n(x) \\ &= \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right)^2 \sqrt{n} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\xi}_{n,\alpha}^2 \sqrt{n} \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \\ &\quad - \sqrt{n} \int_{[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}})} \hat{\mathbb{F}}_n(x) dx^2 \\ &= \left(\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right)^2 - \hat{\xi}_{n,\alpha}^2 \right) \sqrt{n} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) \\ &\quad + \hat{\xi}_{n,\alpha}^2 \sqrt{n} \left(\hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right) \\ &\quad - 2 \int_{[0,z)} \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy \\ &= \left(2z \hat{\xi}_{n,\alpha} + \frac{z^2}{\sqrt{n}} \right) \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) \\ &\quad + \hat{\xi}_{n,\alpha}^2 \sqrt{n} \left(\hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right) \\ &\quad - 2 \left(\hat{\xi}_{n,\alpha} \int_{[0,z)} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy + \int_{[0,z)} \frac{y}{\sqrt{n}} \hat{\mathbb{F}}_n \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy \right) \\ &\xrightarrow{P} 2z \xi_\alpha F(\xi_\alpha) + z \xi_\alpha^2 f(\xi_\alpha) - 2z \xi_\alpha F(\xi_\alpha) = z \xi_\alpha^2 f(\xi_\alpha) \end{aligned}$$

and analogously for $z < 0$. Combining results we establish $I \xrightarrow{P^*} \xi_\alpha z f(\xi_\alpha) \Omega'$ in probability. Consider the second term; since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ (Theorem 4.1) and $\hat{\theta}_n^* \xrightarrow{P^*} \theta_0$ almost surely (Lemma 4.5), we have $\mathbb{P}^* [\check{\theta}_n \notin \mathcal{V}(\theta_0)] \xrightarrow{a.s.} 0$. Thus, for every $\varepsilon > 0$ we obtain

$$\mathbb{P}^* [||II|| \geq \varepsilon]$$

$$\begin{aligned}
 &\leq \mathbb{P}^* \left[\left\| \frac{1}{n} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left(\tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n)\tilde{D}'_t(\check{\theta}_n) \right) j_{n,t}^{*(2)} \right\| \geq \varepsilon \cap \check{\theta}_n \in \mathcal{V}(\theta_0) \right] \\
 &\quad + \mathbb{P}^* [\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\
 &\leq \mathbb{P}^* \left[\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) |j_{n,t}^{*(2)}| \geq \varepsilon \right] \\
 &\quad + o(1) \\
 &\leq \frac{1}{\varepsilon} \mathbb{E}^* \left[\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) |j_{n,t}^{*(2)}| \right] \\
 &\quad + o(1) \\
 &= \frac{1}{\varepsilon} \mathbb{E}^* [|j_{n,t}^{*(2)}|] \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) \\
 &\quad + o(1)
 \end{aligned}$$

almost surely, where the third inequality follows from Markov's inequality. Because $\mathbb{E}^* [|j_{n,t}^{*(2)}|] \leq \mathbb{E}^* [(j_{n,t}^{*(2)})^2]^{\frac{1}{2}} \xrightarrow{p} 0$, it remains to show that

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) \quad (4.95)$$

is stochastically bounded. Using (4.32) we find

$$\begin{aligned}
 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left(\|H_t(\theta)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|H_t(\theta)\|) \right) \\
 &\leq V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t).
 \end{aligned}$$

Employing (4.74) we further have

$$\begin{aligned}
 \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left(\frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} + \frac{C_1 \rho^t}{\underline{\omega}} \left(1 + \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} \right) \right) \\
 &\stackrel{a.s.}{\leq} S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t).
 \end{aligned}$$

In addition, (4.27) and (4.36) imply

$$\begin{aligned}
 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left(\|D_t(\theta)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta)\|) \right)^2 \\
 &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} 3 \left(\|D_t(\theta)\|^2 + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + \|D_t(\theta)\|)^2 \right) \\
 &\leq 3U_t^2 + \frac{3C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + U_t^2).
 \end{aligned}$$

Hence,

$$\begin{aligned}
 &\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\| \right)^2 \\
 &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \left(S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \\
 &\quad \times \left(V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) + 6U_t^2 + \frac{6C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + U_t^2) \right) \\
 &= \underbrace{\frac{1}{n} \sum_{t=1}^n S_t T_t V_t}_{II_1} + \underbrace{\frac{6}{n} \sum_{t=1}^n S_t T_t U_t^2}_{II_2} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t}_{II_3} \\
 &\quad + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t V_t}_{II_4} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t V_t}_{II_5} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{6}{n} \sum_{t=1}^n \rho^t U_t^2}_{II_6} \\
 &\quad + \underbrace{\frac{C_1}{\underline{\omega}} \frac{6}{n} \sum_{t=1}^n \rho^t S_t T_t U_t^2}_{II_7} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t V_t}_{II_8} + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} V_t}_{II_9} \\
 &\quad + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t}_{II_{10}} + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t V_t}_{II_{11}} + \underbrace{\frac{6C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t}_{II_{12}} \\
 &\quad + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} U_t^2}_{II_{13}} + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} S_t T_t}_{II_{14}} + \underbrace{\frac{6C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t U_t^2}_{II_{15}} \\
 &\quad + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} S_t T_t U_t^2}_{II_{16}} + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t}}_{II_{17}} + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t}}_{II_{18}}
 \end{aligned}$$

From Assumption 4.9, the uniform ergodic theorem and Hölder's inequality, we obtain

$$\begin{aligned} II_1 &\leq \left(\frac{1}{n} \sum_{t=1}^n S_t^3 \right)^{\frac{1}{3}} \left(\frac{1}{n} \sum_{t=1}^n T_t^3 \right)^{\frac{1}{3}} \left(\frac{1}{n} \sum_{t=1}^n V_t^3 \right)^{\frac{1}{3}} \\ &\xrightarrow{a.s.} \left(\mathbb{E}[S_t^3] \right)^{\frac{1}{3}} \left(\mathbb{E}[T_t^3] \right)^{\frac{1}{3}} \left(\mathbb{E}[V_t^3] \right)^{\frac{1}{3}} < \infty \end{aligned}$$

and similarly we can show that $\lim_{n \rightarrow \infty} II_2 < \infty$ almost surely. Consider II_3 ; for each $\varepsilon > 0$, Markov's inequality and the Cauchy-Schwarz inequality yield

$$\sum_{t=1}^{\infty} \mathbb{P} \left[\rho^t S_t T_t > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[S_t T_t]}{\varepsilon} = \frac{1 + (\mathbb{E}[S_t^2])^{\frac{1}{2}} (\mathbb{E}[T_t^2])^{\frac{1}{2}}}{\varepsilon(1 - \rho)} < \infty$$

and $\frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t \xrightarrow{a.s.} 0$ follows from combining the Borel-Cantelli lemma with Cesàro's lemma. Thus, $II_3 \xrightarrow{a.s.} 0$. Similarly we can show that the terms II_4, \dots, II_{16} vanish almost surely. Further, $II_{17} \leq \frac{1}{n} \frac{C_1^2}{\varpi^2(1-\rho^2)} \xrightarrow{a.s.} 0$ and similarly, we can prove that II_{18} vanishes almost surely, which completes the proof. \square

4.B Recursive-design Residual Bootstrap

This appendix devotes attention to the recursive-design residual bootstrap. The bootstrap schemes described in Algorithms 4.3 and 4.4 are the recursive-design counterparts of Algorithms 4.1 and 4.2, respectively. Note that the bootstrap observation ϵ_t^* is generated recursively on the basis of its past realizations $\epsilon_{t-1}^*, \dots, \epsilon_1^*$.

Algorithm 4.3. (*Recursive-design residual bootstrap*)

1. For $t = 1, \dots, n$ generate $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$ and the bootstrap observation $\epsilon_t^* = \sigma_t^* \eta_t^*$ with $\sigma_t^* = \sigma_t^*(\hat{\theta}_n)$ and $\sigma_t^*(\theta) = \sigma(\epsilon_{t-1}^*, \dots, \epsilon_1^*, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta)$
2. Calculate the bootstrap estimator

$$\hat{\theta}_n^* = \arg \max_{\theta \in \Theta} L_n^*(\theta)$$

with the bootstrap criterion function given by

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t^*(\theta) \quad \text{and} \quad \ell_t^*(\theta) = -\frac{1}{2} \left(\frac{\epsilon_t^*}{\sigma_t^*(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

3. For $t = 1, \dots, n$ compute the bootstrap residual $\hat{\eta}_t^* = \epsilon_t^* / \sigma_t^*(\hat{\theta}_n^*)$ and obtain

$$\hat{\xi}_{n,\alpha}^* = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z).$$

4. Obtain the bootstrap estimator of the conditional VaR

$$\widehat{VaR}_{n,\alpha}^* = -\hat{\xi}_{n,\alpha}^* \tilde{\sigma}_{n+1}(\hat{\theta}_n^*).$$

Algorithm 4.4. (Recursive-design bootstrap confidence intervals for VaR)

1. Acquire a set of B bootstrap replicates, i.e. $\widehat{VaR}_{n,\alpha}^{*(b)}$ for $b = 1, \dots, B$, by repeating Algorithm 4.3.

2. (i) Obtain the EP interval

$$\left[\widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2), \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2) \right]$$

$$\text{with } \hat{G}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\left\{ \sqrt{n}(\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}) \leq x \right\}}.$$

(ii) Calculate the RT interval

$$\left[\widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2) \right].$$

(iii) Compute the SY interval

$$\left[\widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma) \right]$$

$$\text{with } \hat{H}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\left\{ \sqrt{n}|\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}| \leq x \right\}}.$$

In contrast to the fixed-design residual bootstrap, the bootstrap sample $\epsilon_1^*, \dots, \epsilon_n^*$, conditional on the original sample, is a dependent sequence. Therefore one likely needs a stronger set of conditions to show the validity of the recursive-design residual bootstrap. Moreover, whether the recursive bootstrap scheme is valid is contingent on the specific conditional volatility model, e.g. GARCH(1, 1), and as such needs to be investigated on a case-by-case basis. This is therefore outside the scope of the present work.

Chapter 5

A Residual Bootstrap for Conditional Expected Shortfall

This chapter studies a fixed-design residual bootstrap method for the two-step estimator of Francq and Zakoïan (2015) associated with the conditional Expected Shortfall. For a general class of volatility models the bootstrap is shown to be asymptotically valid under the conditions imposed in Chapter 4. A simulation study is conducted revealing that the average coverage rates are satisfactory for most settings considered. There is no clear evidence to have a preference for any of the three proposed bootstrap intervals. This contrasts results in Chapter 4 for the VaR, for which the reversed-tails interval has a superior performance.¹

¹This chapter is based on the paper Heinemann and Telg (2018).

5.1 Introduction

The assessment of market risk is a key challenge that financial market participants face on a daily basis. To evaluate the risk, financial institutions primarily employ the risk measure Value-at-Risk (VaR) to meet the capital requirements enforced by the Basel Committee on Banking Supervision. Despite its popularity, the VaR is not a coherent risk measure as it fails to fulfill the subadditivity property (Artzner et al., 1999). A coherent alternative is the related risk measure Expected Shortfall (ES). For a given level α , it is defined as the expected return in the worst $100\alpha\%$ cases and is therefore sometimes called Expected Tail Loss.² In contrast to the VaR, the ES provides valuable information on the severity of an incurred loss, which makes it the preferred risk measure in practice (cf. Acerbi and Tasche, 2002a; 2002b). Consequently, the Basel Committee published revised standards in January 2016 resembling a shift from VaR towards ES as the underlying risk measure (Osmundsen, 2018).

In the literature there is an increasing interest in conditional risk measures, which take into account the temporal dependence of asset returns. Frequently, the volatility dynamics are specified by a (semi-)parametric model such that the conditional ES can be expressed as the product of the conditional volatility and the ES of the innovations distribution (cf. Francq and Zakoïan, 2015, Example 2). The latter can be treated as additional parameter, which is generally unknown just like the parameters of the conditional volatility model. Inferring the parameters from data leads to an evaluation of the conditional ES that is prone to estimation risk. As argued in Chapter 4 this estimation uncertainty can be substantial for risk measures related to extreme events.

The uncertainty around point estimates is typically determined by asymptotic theory, in which one replaces unknown quantities in the limiting distribution by consistent estimates. For example, Cai and Wang (2008) and Martins-Filho, Yao, and Torero (2018) study the behavior of proposed nonparametric estimators for conditional VaR and ES, based on asymptotics and simulation studies. An alternative approach is based on bootstrap approximations. Regarding the estimators of the volatility models parameters, several bootstrap methods have been examined, among which the sub-sample bootstrap (Hall and Yao, 2003), the block bootstrap

²In the literature, ES is also sometimes referred to as conditional VaR since it is defined as the expected loss given a VaR exceedence. Since conditional refers to temporal dependence (i.e. conditional on past returns) in this work, we refrain from using this term to prevent any confusion.

(Corradi and Iglesias, 2008), the wild bootstrap (Shimizu, 2010) and the residual bootstrap, both with recursive (Pascual et al., 2006; Hidalgo and Zaffaroni, 2007) and fixed design (Shimizu, 2010; Cavaliere et al., 2018). However, the estimation of the conditional ES has received only limited attention in the bootstrap literature. Christoffersen and Gonçalves (2005) construct intervals for conditional ES based on a recursive-design residual bootstrap method. Gao and Song (2008) compare coverage probabilities for conditional ES based on this bootstrap method and asymptotic normality results in their simulation study.

In this chapter, we extend the results of Chapter 4 derived for conditional VaR to the conditional ES estimator. In particular, we follow the two-step procedure of Francq and Zakoïan (2015) for the estimation of the underlying parameters. In a first step, we obtain estimates of the parameters of the stochastic volatility model by quasi-maximum-likelihood (QML) estimation. Based on the model's residuals, an estimate for the innovations ES is obtained in the second step. We propose a fixed-design residual bootstrap method to mimic the finite sample distribution of this two-step estimator for a general class of volatility models. Moreover, an algorithm is provided for the construction of bootstrap intervals for the conditional ES.

The remainder of the chapter is organized as follows. Section 5.2 introduces a general class of volatility models and derives the conditional ES. The two-step estimation procedure is described in Section 5.3 and corresponding asymptotic results are provided under the assumptions imposed in Chapter 4. In Section 5.4, a fixed-design residual bootstrap method is proposed and proven to be consistent. In addition, bootstrap intervals are constructed for the conditional ES. Section 5.5 consists of a Monte Carlo study. Section 5.6 concludes. Auxiliary results and proofs of the main results are gathered in the Appendix.

5.2 Model

We consider conditional volatility models of the form

$$\epsilon_t = \sigma_t \eta_t, \tag{5.1}$$

with $t \in \mathbb{Z}$, where ϵ_t denotes the log-return, $\{\sigma_t\}$ is a volatility process and $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) variables. The

volatility is assumed to be a measurable function of past observations

$$\sigma_{t+1} = \sigma_{t+1}(\theta_0) = \sigma(\epsilon_t, \epsilon_{t-1}, \dots; \theta_0), \quad (5.2)$$

with $\sigma : \mathbb{R}^\infty \times \Theta \rightarrow (0, \infty)$ and θ_0 denotes the true parameter vector belonging to the parameter space $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}$. Various commonly used volatility models satisfy (5.1) – (5.2); for examples see Francq and Zakoïan (2015, Table 1). Consider an arbitrary real-valued random variable X (e.g. stock return) with cdf F_X . If $\mathbb{E}_X[X^-] < \infty$ with $X^- = \max\{-X, 0\}$, then the ES at level $\alpha \in (0, 1)$ is finite and given by $ES_\alpha(X) = -\mathbb{E}_X[X|X < F_X^{-1}(\alpha)]$. Let \mathcal{F}_{t-1} denote the σ -algebra generated by $\{\epsilon_u, u < t\}$. It follows that the conditional ES of ϵ_t given \mathcal{F}_{t-1} at level $\alpha \in (0, 1)$ is

$$ES_\alpha(\epsilon_t|\mathcal{F}_{t-1}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) ES_\alpha(\eta_t). \quad (5.3)$$

As η_t are i.i.d., the ES at level α of η_t is constant for a given α and can be treated as a parameter. Setting $\mu_\alpha = -\mathbb{E}[\eta_t|\eta_t < \xi_\alpha]$ with $\xi_\alpha = F^{-1}(\alpha)$ and F denoting the cdf of η_t , (5.3) reduces to

$$ES_\alpha(\epsilon_t|\mathcal{F}_{t-1}) = \mu_\alpha \sigma_t(\theta_0). \quad (5.4)$$

Typically, α is chosen small (e.g. 5% or 1%) such that $\xi_\alpha < 0$ and hence $\mu_\alpha > 0$. Except for special cases³, μ_α is unknown and needs to be estimated just like θ_0 .

5.3 Estimation

For the estimation of the parameters θ_0 and μ_α we employ the two-step procedure of Francq and Zakoïan (2015, Remark 3). First, the vector of the conditional volatility parameters θ_0 is estimated by quasi-maximum-likelihood (QML). Since

$$\sigma_{t+1}(\theta) = \sigma(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta) \quad (5.5)$$

³We derive the analytical expressions for μ_α for the cases in which η_t are normally as well as Student- t distributed in Appendix 5.B.

can generally not be determined completely given a sample $\epsilon_1, \dots, \epsilon_n$, we replace the unknown presample observations by arbitrary values, say $\tilde{\epsilon}_t$, $t \leq 0$, yielding

$$\tilde{\sigma}_{t+1}(\theta) = \sigma(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta). \quad (5.6)$$

Then the QML estimator of θ_0 is defined as

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta) \quad (5.7)$$

with the criterion function specified by

$$\tilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\theta) \quad \text{and} \quad \tilde{\ell}_t(\theta) = -\frac{1}{2} \left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

In the second step, μ_α can be estimated on the basis of the first-step residuals, i.e. $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$. A reasonable estimator of μ_α (cf. Gao and Song, 2008) is given by

$$\hat{\mu}_{n,\alpha} = -\frac{\sum_{t=1}^n \hat{\eta}_t \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}}}{\sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}}}, \quad (5.8)$$

where $\hat{\xi}_{n,\alpha}$ is the empirical α -quantile of $\hat{\eta}_1, \dots, \hat{\eta}_n$, i.e. $\hat{\xi}_{n,\alpha} = \hat{\mathbb{F}}_n^{-1}(\alpha)$ with empirical distribution function $\hat{\mathbb{F}}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}}$.

Having obtained estimators for θ_0 and μ_α , we turn to the estimation of the conditional ES of the one-period ahead observation at level α . For notational convenience, we use the abbreviation $ES_{n,\alpha}$ to denote $ES_\alpha(\epsilon_{n+1} | \mathcal{F}_n)$. Employing (5.6) – (5.8) we can estimate $ES_{n,\alpha}$ by

$$\widehat{ES}_{n,\alpha} = \hat{\mu}_{n,\alpha} \tilde{\sigma}_{n+1}(\hat{\theta}_n). \quad (5.9)$$

For the asymptotic analysis of (5.7) – (5.9) we assume the conditions of Chapter 4, which we restate for completeness.

Assumption 5.1. (*Compactness*) Θ is a compact subset of \mathbb{R}^r .

Assumption 5.2. (*Stationarity & ergodicity*) $\{\epsilon_t\}$ is a strictly stationary and ergodic solution of (5.1) with (5.2).

Assumption 5.3. (*Volatility process*) For any real sequence $\{x_i\}$, the function $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$ is continuous. Almost surely, $\sigma_t(\theta) > \underline{\omega}$ for any $\theta \in \Theta$ and some $\underline{\omega} > 0$ and $\mathbb{E}[\sigma_t^s(\theta_0)] < \infty$ for some $s > 0$. Moreover, for any $\theta \in \Theta$, we

assume $\sigma_t(\theta_0)/\sigma_t(\theta) = 1$ almost surely (a.s.) if and only if $\theta = \theta_0$.

Assumption 5.4. (*Initial conditions*) There exists a constant $\rho \in (0, 1)$ and a random variable C_1 measurable with respect to \mathcal{F}_0 and $\mathbb{E}[C_1^s] < \infty$ for some $s > 0$ such that

$$(i) \sup_{\theta \in \Theta} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \leq C_1 \rho^t;$$

(ii) $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives satisfying

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} \right\| \leq C_1 \rho^t, \quad \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho^t,$$

where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 5.5. (*Innovation process*) The innovations $\{\eta_t\}$ satisfy

(i) $\eta_t \stackrel{iid}{\sim} F$ with F being continuous, $\mathbb{E}[\eta_t^2] = 1$ and η_t is independent of $\{\epsilon_u : u < t\}$;

(ii) η_t admits a density f which is continuous and strictly positive around $\xi_\alpha < 0$;

(iii) $\mathbb{E}[\eta_t^4] < \infty$.

Assumption 5.6. (*Interior*) θ_0 belongs to the interior of Θ denoted by $\mathring{\Theta}$.

Assumption 5.7. (*Non-degeneracy*) There does not exist a non-zero $\lambda \in \mathbb{R}^r$ such that $\lambda' \frac{\partial \sigma_t(\theta_0)}{\partial \theta} = 0$ almost surely.

Assumption 5.8. (*Monotonicity*) For any real sequence $\{x_i\}$ and for any $\theta_1, \theta_2 \in \Theta$ satisfying $\theta_1 \leq \theta_2$ componentwise, we have $\sigma(x_1, x_2, \dots; \theta_1) \leq \sigma(x_1, x_2, \dots; \theta_2)$.

Assumption 5.9. (*Moments*) There exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that the following variables have finite expectation:

$$(i) \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^a,$$

$$(ii) \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^b,$$

$$(iii) \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^c$$

for some a, b, c (to be specified).

Assumption 5.10. (*Scaling stability*) There exists a function g such that for any $\theta \in \Theta$, for any $\lambda > 0$, and any real sequence $\{x_i\}$

$$\lambda\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; \theta_\lambda),$$

where $\theta_\lambda = g(\theta, \lambda)$ and g is differentiable in λ .

For a discussion of the conditions we refer to Section 4.3 and Francq and Zakoïan (2015, Section 2 and 3). On the basis of the previous assumptions we extend the strong consistency result of Francq and Zakoïan (2015, Theorem 1) to the estimator of the ES at level α of η_t .

Theorem 5.1. (*Strong consistency*) Under Assumptions 5.1–5.3, 5.4(i) and 5.5(i) the estimator in (5.7) is strongly consistent, i.e. $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$. If in addition Assumptions 5.5(iii), 5.6, and 5.9(i) hold with $a = -1, 4$, then the estimator in (5.8) satisfies $\hat{\mu}_{n,\alpha} \xrightarrow{a.s.} \mu_\alpha$.

Proof. Francq and Zakoïan (2015, Theorem 1) establish $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$. Moreover, we have

$$\hat{\mu}_{n,\alpha} = -\frac{\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}}}{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}}} \xrightarrow{a.s.} -\frac{\mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]}{\mathbb{P}[\eta_t < \xi_\alpha]} = -\mathbb{E}[\eta_t | \eta_t < \xi_\alpha] = \mu_\alpha$$

by Lemma 4.2, which verifies the second claim. \square

To lighten notation, we henceforth write $D_t(\theta) = \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}$ and drop the argument when evaluated at the true parameter, i.e. $D_t = D_t(\theta_0)$. The next result provides the joint asymptotic distribution of $\hat{\theta}_n$ and $\hat{\mu}_{n,\alpha}$ and is due to Francq and Zakoïan (2012).

Theorem 5.2. (*Asymptotic distribution*) Suppose Assumptions 5.1–5.7, 5.9 and 5.10 hold with $a = b = 4$ and $c = 2$. Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \sqrt{n}(\hat{\mu}_{n,\alpha} - \mu_\alpha) \end{pmatrix} \xrightarrow{d} N(0, \Gamma_\alpha) \quad \text{with} \quad \Gamma_\alpha = \begin{pmatrix} \frac{\kappa-1}{4} J^{-1} & \varphi_\alpha J^{-1} \Omega \\ \varphi_\alpha \Omega' J^{-1} & \nu_\alpha \end{pmatrix}, \quad (5.10)$$

where $\kappa = \mathbb{E}[\eta_t^4]$, $\Omega = \mathbb{E}[D_t]$, $J = \mathbb{E}[D_t D_t']$, $\varphi_\alpha = \frac{1}{2}x_\alpha - \mu_\alpha \frac{\kappa-1}{4}$, $\nu_\alpha = \sigma_\alpha^2 - x_\alpha \mu_\alpha + \frac{\kappa-1}{4} \mu_\alpha^2$, $\sigma_\alpha^2 = \frac{1}{\alpha^2} \text{Var}[(\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]$ and $x_\alpha = -\frac{1}{\alpha} \text{Cov}[\eta_t^2, (\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]$.

In order to evaluate σ_α^2 and x_α in Theorem 5.2, we need expressions for the variance and covariance term respectively. After basic manipulation we find

$$\begin{aligned}\text{Var}[(\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] &= p_\alpha + \alpha + \xi_\alpha(1 - \alpha)\alpha(\xi_\alpha - 2\mu_\alpha) - (\alpha\mu_\alpha)^2 \\ \text{Cov}[\eta_t^2, (\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] &= -\alpha\mu_\alpha - (\xi_\alpha p_\alpha - q_\alpha),\end{aligned}$$

with $p_\alpha = \mathbb{E}[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \alpha$ and $q_\alpha = \mathbb{E}[\eta_t^3 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]$. In a GARCH(p, q) setting, Gao and Song (2008) quantify the uncertainty around $\hat{\theta}_n$ and $\hat{\mu}_{n,\alpha}$ using (5.10) while replacing the unknown quantities in Γ_α by estimates. In the same spirit, ξ_α and μ_α can be substituted by $\hat{\xi}_{n,\alpha}$ and $\hat{\mu}_{n,\alpha}$ while Ω , J , q_α , p_α and κ can be replaced by

$$\begin{aligned}\hat{\Omega}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t, & \hat{q}_{n,\alpha} &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^3 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}}, & \hat{\kappa}_n &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^4, \\ \hat{J}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t', & \hat{p}_{n,\alpha} &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} - \alpha,\end{aligned}\tag{5.11}$$

where $\hat{D}_t = \tilde{D}_t(\hat{\theta}_n)$ and $\tilde{D}_t(\theta) = \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta}$. The strong consistency of the estimators in (5.11) follows from Lemma 4.2 and Theorem 4.1. Based on (5.11) we obtain a consistent estimator for Γ_α denoted by $\hat{\Gamma}_{n,\alpha}$. Note that in the joint asymptotic distribution in Theorem 5.2 the probability density function (pdf) of η_t does not occur. This is in contrast to the limiting distribution of the parameters that comprise the conditional VaR estimator (see (4.10)). Hence, no density estimation (by e.g. kernel smoothing) is required here.

The asymptotic behavior of the conditional ES estimator can be studied by employing Theorem 5.2. Since the conditional volatility varies over time, a limiting distribution cannot exist and therefore the concept of weak convergence is not applicable in this context. In Chapter 2 we study a *merging* concept that generalizes the notion of weak convergence, i.e. two sequences of (random) cdfs $\{F_n\}, \{G_n\}$ *merge* (in probability) if and only if their bounded Lipschitz distance $d_{BL}(F_n, G_n)$ converges to zero (in probability). Assuming two independent samples, one for parameter estimation and one for conditioning, the delta method suggests that the ES estimator, centered at $ES_{n,\alpha}$ and inflated by \sqrt{n} , and

$$N\left(0, \left(\mu_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta}\right)' \Gamma_\alpha \left(\mu_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta}\right)\right)\tag{5.12}$$

given \mathcal{F}_n merge in probability. Equation (5.12) highlights once more the relevance of the merging concept since the conditional variance still depends on n and does not converge as $n \rightarrow \infty$. In combination with Theorem 5.1 and $\hat{\Gamma}_{n,\alpha} \xrightarrow{a.s.} \Gamma_\alpha$, $100(1 - \gamma)\%$ confidence intervals for $ES_{n,\alpha}$ can be constructed with bounds given by

$$\widehat{ES}_{n,\alpha} \pm \frac{\Phi^{-1}(\gamma/2)}{\sqrt{n}} \left\{ \begin{pmatrix} \hat{\mu}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix}' \hat{\Gamma}_{n,\alpha} \begin{pmatrix} \hat{\mu}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix} \right\}^{\frac{1}{2}}, \quad (5.13)$$

where Φ denotes the standard normal cdf. It has to be mentioned that researchers rarely have a replicate, independent of the original series, to their disposal.⁴ An asymptotic justification for the interval on the basis of a single sample is given in Chapter 2. Bootstrap methods offer an alternative way to quantify the uncertainty around the estimators.

5.4 Bootstrap

5.4.1 Fixed-design Residual Bootstrap

We propose a fixed-design residual bootstrap procedure, described in Algorithm 5.1, to approximate the distribution of the estimators in (5.7) – (5.9).

Algorithm 5.1. (*Fixed-design residual bootstrap*)

1. For $t = 1, \dots, n$, generate $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$ and the bootstrap observation $\epsilon_t^* = \tilde{\sigma}_t(\hat{\theta}_n)\eta_t^*$, where $\tilde{\sigma}_t(\theta)$ and $\hat{\theta}_n$ are given in (5.6) and (5.7), respectively.
2. Calculate the bootstrap estimator

$$\hat{\theta}_n^* = \arg \max_{\theta \in \Theta} L_n^*(\theta) \quad (5.14)$$

with the bootstrap criterion function given by

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t^*(\theta) \quad \text{and} \quad \ell_t^*(\theta) = -\frac{1}{2} \left(\frac{\epsilon_t^*}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

⁴Exceptions would include some experimental settings.

3. For $t = 1, \dots, n$ compute the bootstrap residual $\hat{\eta}_t^* = \epsilon_t^* / \tilde{\sigma}_t(\hat{\theta}_n^*)$ and obtain

$$\hat{\mu}_{n,\alpha}^* = - \frac{\sum_{t=1}^n \hat{\eta}_t^* \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}}}{\sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}}}, \quad (5.15)$$

where $\hat{\xi}_{n,\alpha}^*$ is the empirical α -quantile of $\hat{\eta}_1^*, \dots, \hat{\eta}_n^*$.

4. Obtain the bootstrap estimator of the conditional ES

$$\widehat{ES}_{n,\alpha}^* = \hat{\mu}_{n,\alpha}^* \tilde{\sigma}_{n+1}(\hat{\theta}_n^*). \quad (5.16)$$

In the following subsection we show the asymptotic validity of the fixed-design bootstrap procedure described in Algorithm 5.1.

5.4.2 Bootstrap Consistency

Subsequently, we employ the usual notation for bootstrap asymptotics, i.e. “ $\xrightarrow{p^*}$ ” and “ $\xrightarrow{d^*}$ ”, as well as the standard bootstrap stochastic order symbol “ $o_{p^*}(1)$ ” (cf. Chang and Park, 2003). The asymptotic validity of the bootstrap corresponding to the stochastic volatility part is shown in Proposition 4.1. Therefore, we focus only on $\hat{\mu}_{n,\alpha}^*$. By construction, we have $\sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} = \lfloor \alpha n \rfloor + 1$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x . Defining $\alpha_n = \frac{\lfloor \alpha n \rfloor + 1}{n}$, we standardize (5.15) such that the bootstrap estimator satisfies

$$\begin{aligned} \sqrt{n}(\hat{\mu}_{n,\alpha}^* - \hat{\mu}_{n,\alpha}) &= - \frac{1}{\alpha_n} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\hat{\eta}_t^* \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} + \alpha_n \hat{\mu}_{n,\alpha} \right) \\ &= - \frac{1}{\alpha_n} (A_n^* + B_n^* + C_n^* + D_n^*), \end{aligned} \quad (5.17)$$

where the scaling factor $-1/\alpha_n$ in (5.17) converges to $-1/\alpha$ since $\alpha \leq \alpha_n \leq \alpha + \frac{1}{n}$. The different terms in brackets are given by

$$\begin{aligned} A_n^* &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}) (\mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}), \\ B_n^* &= \hat{\xi}_{n,\alpha} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \alpha_n), \\ C_n^* &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \eta_t^*) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}, \end{aligned}$$

$$D_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left((\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} + \alpha_n (\hat{\xi}_{n,\alpha} + \hat{\mu}_{n,\alpha}) \right).$$

Employing arguments of Chen (2008, Lemma 2) Lemma 5.1 in Appendix 5.A states the asymptotic negligibility of A_n^* , i.e. $A_n^* \xrightarrow{p^*} 0$ in probability. The term $B_n^* = 0$ since $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} = \alpha_n$ by construction. Further, Lemma 5.2 in Appendix 5.A states that $C_n^* = \alpha \mu_\alpha \Omega \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$ in probability. Last, we have $D_n^* \xrightarrow{d^*} N(0, \nu_\alpha)$ almost surely by Lemma 5.3 in Appendix 5.A. The previous discussion together with the asymptotic expansion of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ in (4.18) yields

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\mu}_{n,\alpha}^* - \hat{\mu}_{n,\alpha}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} J^{-1} & O_{r \times 1} \\ -\frac{1}{2} \mu_\alpha \Omega' J^{-1} & -\frac{1}{\alpha} \end{pmatrix} \\ \times \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t (\eta_t^{*2} - 1) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \left((\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} + \alpha_n (\hat{\xi}_{n,\alpha} + \hat{\mu}_{n,\alpha}) \right) \end{pmatrix} + o_{p^*}(1)$$

in probability. Employing Lemma 5.3 once more leads to the chapter's main result.

Theorem 5.3. (*Bootstrap consistency*) Suppose Assumptions 5.1–5.10 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\mu}_{n,\alpha}^* - \hat{\mu}_{n,\alpha}) \end{pmatrix} \xrightarrow{d^*} N(0, \Gamma_\alpha)$$

in probability.

Theorem 5.3 is useful to validate the bootstrap for the conditional ES estimator. For the asymptotic behavior of the conditional ES estimator we refer to (5.12) and the text preceding it. The following corollary is established.

Corollary 5.1. Under the assumptions of Theorem 5.3 the conditional cdfs of $\sqrt{n}(\widehat{ES}_{n,\alpha}^* - \widehat{ES}_{n,\alpha})$ given \mathcal{F}_n and (5.12) given \mathcal{F}_n merge in probability.

Having proven first-order asymptotic validity of the bootstrap procedure described in Algorithm 5.1, we turn to constructing bootstrap intervals for ES.

5.4.3 Bootstrap Confidence Intervals for ES

Clearly, the ES evaluation in (5.9) is subject to estimation risk that needs to be quantified. We propose the following algorithm to obtain approximately $100(1 - \gamma)\%$ confidence intervals.

Algorithm 5.2. (*Fixed-design bootstrap confidence intervals for ES*)

1. Acquire a set of B bootstrap replicates, i.e. $\widehat{ES}_{n,\alpha}^{*(b)}$ for $b = 1, \dots, B$, by repeating Algorithm 5.1.
2. (i) Obtain the equal-tailed percentile (EP) interval

$$\left[\widehat{ES}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2), \widehat{ES}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2) \right] \quad (5.18)$$

$$\text{with } \hat{G}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\left\{ \sqrt{n}(\widehat{ES}_{n,\alpha}^{*(b)} - \widehat{ES}_{n,\alpha}) \leq x \right\}}.$$

- (ii) Calculate the reversed-tails (RT) interval

$$\left[\widehat{ES}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2), \widehat{ES}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2) \right]. \quad (5.19)$$

- (iii) Compute the symmetric (SY) interval

$$\left[\widehat{ES}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma), \widehat{ES}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma) \right] \quad (5.20)$$

$$\text{with } \hat{H}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\left\{ \sqrt{n}|\widehat{ES}_{n,\alpha}^{*(b)} - \widehat{ES}_{n,\alpha}| \leq x \right\}}.$$

For a discussion of the three interval types in Algorithm 5.2, we refer to Section 4.4.3. In the next section, features of the fixed-design bootstrap intervals for the conditional ES are studied by means of simulations.

5.5 Monte Carlo Experiment

To assess the proposed bootstrap procedure in finite samples, we consider a simulation setup similar to Chapter 4. The Data Generating Process (DGP) is a GARCH(1, 1), which falls in the class of conditional volatility models defined in

(5.1) – (5.2). More specifically, we consider

$$\begin{cases} \epsilon_t = \sigma_t \eta_t, \\ \sigma_{t+1}^2 = \omega_0 + \alpha_0 \epsilon_t^2 + \beta_0 \sigma_t^2 \end{cases}$$

with $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. Regarding the GARCH parameters we study two scenarios:

- (i) high persistence: $\theta_0 = (0.05 \times 20^2/252, 0.15, 0.8)$,
- (ii) low persistence: $\theta_0 = (0.05 \times 20^2/252, 0.4, 0.55)$.

The innovations $\{\eta_t\}$ are drawn from two different distributions: the Student- t distribution with $\nu = 6$ degrees of freedom and the standard normal distribution (which corresponds to the case $\nu = \infty$). Whereas in the latter case the innovations are appropriately standardized, in the former we draw from the normalized density $f(x) = \frac{1}{\sigma_\nu} f_\nu(x/\sigma_\nu)$ such that $\mathbb{E}[\eta_t^2] = 1$, where $\sigma_\nu^2 = \frac{\nu-2}{\nu}$ and $f_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} (1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$. In this setting, the ES of the innovations distribution reduces to $\mu_\alpha = \frac{f_{\nu-2}(\xi_\alpha)}{\alpha}$ with $\xi_\alpha = \sigma_\nu F_\nu^{-1}(\alpha)$ and $F_\nu(x) = \int_{-\infty}^x f_\nu(y)dy$; we refer to Appendix 5.B for details. For the experiment, the ES level takes two values: $\alpha \in \{0.01, 0.05\}$. We consider four different sample sizes $n \in \{500; 1,000; 5,000; 10,000\}$ and the number of bootstrap replicates is fixed at $B = 2,000$. For each model, we simulate $S = 2,000$ independent Monte Carlo trajectories. All simulations are carried out on a HP Z640 workstation with 16 cores using Matlab R2016a. The numerical optimization of the log-likelihood function is performed using the built-in function *fmincon*. Parallel computing by means of *parfor* is employed to reduce running time significantly.

In Section 4.5 it is demonstrated that the bootstrap distribution mimics adequately the finite sample distribution of the estimator of the volatility parameters. In a similar fashion, we assess whether the bootstrap distribution (given a particular sample) mimics the distribution of the ES parameter estimator. Figure 5.1 displays the density estimates for the distribution of $\sqrt{n}(\hat{\mu}_{n,\alpha} - \mu_\alpha)$ and $\sqrt{n}(\hat{\mu}_{n,\alpha}^* - \hat{\mu}_{n,\alpha})$ in the high persistence case for $n = 5,000$ with $\alpha \in \{0.01, 0.05\}$. In both cases, we observe that the density plots are bell curves around the value zero, which supports the theoretical results of Theorem 5.2 and 5.3. Since the density graphs for the other scenarios are very similar, they are not reported in order to conserve space. We continue by studying the coverage probabilities of the three bootstrap intervals introduced in Section 5.4.3.

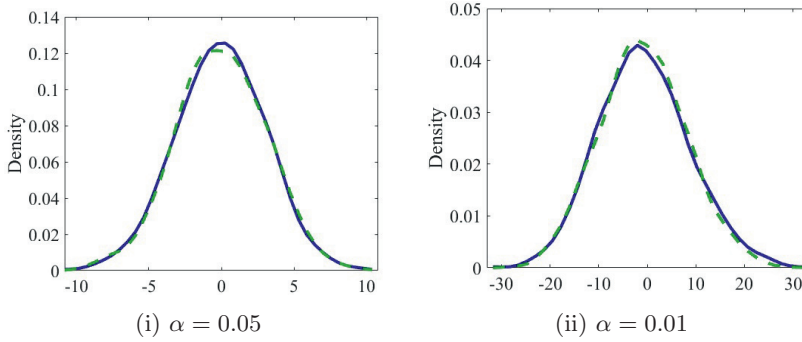


Figure 5.1: Density estimates for the distribution of $\sqrt{n}(\hat{\mu}_{n,\alpha} - \mu_\alpha)$ (full line) based on $S = 2,000$ simulations and the fixed-design bootstrap distribution of $\sqrt{n}(\hat{\mu}_{n,\alpha}^* - \hat{\mu}_{n,\alpha})$ (dashed line) based on $B = 2,000$ replications. The DGP is a GARCH(1, 1) with $\theta_0 = (0.08, 0.15, 0.8)'$, sample size $n = 5,000$ and (normalized) Student- t innovations (6 degrees of freedom).

Table 5.1 reports the results of the three 90%-bootstrap intervals for the 5%-ES with Student- t distributed innovations (which we consider as benchmark). For moderate sample sizes, we observe satisfactory coverage probabilities that lie relatively close to the nominal level of 90%. For small sample size ($n = 500$), the intervals exhibit small under-coverage with values ranging from 4.00 to 5.85 percentage points (pp) below the nominal value. For all three intervals, we find that the average rate of the conditional ES being below the interval is considerably less than it being above the interval. This phenomenon is most pronounced in smaller sample size. Concerning the average length of the intervals, we can make two important observations. Firstly, the SY interval is on average larger than the EP/RT interval.⁵ As sample size increases, this gap disappears and all intervals' average lengths shrink. Secondly, the average length of intervals is larger in the high persistence case, as the conditional volatility varies more compared to the lower persistence case. In the following, we study deviations from the benchmark specification. Table 5.2 considers a change in the innovation distribution F , while Table 5.3 and 5.4 take into account a change in the ES level α and a change in the nominal coverage probability $100(1 - \gamma)\%$, respectively.

Table 5.2 considers the case where the innovations follow a standard normal distribution. Results are qualitatively similar to the benchmark. In particular, coverage rates are generally close to the nominal level for $n \geq 1,000$ yet the under-

⁵By construction, the EP and the RT interval are of equal length.

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
		low persistence			high persistence		
500	EP	84.35	3.85/11.80	0.613	83.15	4.45/12.40	0.825
	RT	85.30	2.35/12.35	0.613	86.00	2.15/11.85	0.825
	SY	86.00	2.50/11.50	0.623	85.45	2.65/11.90	0.833
1,000	EP	88.05	3.40/8.55	0.448	87.55	3.95/8.50	0.610
	RT	88.70	2.70/8.60	0.448	88.85	2.75/8.40	0.610
	SY	88.85	2.60/8.55	0.452	88.55	3.00/8.45	0.612
5,000	EP	89.35	4.35/6.30	0.216	89.85	3.35/6.80	0.287
	RT	89.90	3.75/6.35	0.216	89.90	3.45/6.65	0.287
	SY	90.05	4.05/5.90	0.216	90.15	3.20/6.65	0.287
10,000	EP	89.00	4.70/6.30	0.155	89.60	4.35/6.05	0.204
	RT	89.60	4.55/5.85	0.155	89.55	4.55/5.90	0.204
	SY	89.55	4.35/6.10	0.155	89.55	4.45/6.00	0.204

Table 5.1: The table reports distinct features of the fixed-design bootstrap confidence intervals for the conditional ES at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes, the interval's average coverage rates (in %), the average rate of the conditional ES being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. The DGP is a GARCH(1,1) with (normalized) **Student- t innovations (6 degrees of freedom)**.

coverage in smaller sample sizes is less in this scenario. For example, the average coverage is at most 3.70pp below the 90% level even when $n = 500$. In general, results seem to be “less extreme” compared to the benchmark: (i) the average length of all intervals is smaller for all sample sizes and (ii) the average rate of the conditional ES being above the interval lies closer to the corresponding rate below the interval. Moreover, we observe that there is no interval that outperforms the others in the case of η_t being standard normally distributed.

Table 5.3 provides simulation results for the conditional ES at level $\alpha = 0.01$, where the DGP is a GARCH(1,1) with Student- t innovations (6 degrees of freedom). Unsurprisingly, we find that the average length of all intervals is considerably larger compared to the benchmark. More strikingly, we observe that the phenomenon of under-coverage appears across sample sizes. For the lowest sample size considered, i.e. $n = 500$, average coverage rates are between 15pp and 20pp below nominal value. This problem is still severe for the case $n = 1,000$, as rates are still approximately 10pp too low. Results are more satisfactory for the two

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
		low persistence			high persistence		
500	EP	86.45	4.65/8.90	0.446	86.30	5.10/8.60	0.549
	RT	87.85	2.15/10.00	0.446	87.65	2.40/9.95	0.549
	SY	88.10	2.80/9.10	0.455	87.50	3.50/9.00	0.555
1,000	EP	89.05	4.60/6.35	0.309	89.10	4.25/6.65	0.397
	RT	89.25	3.05/7.70	0.309	88.90	3.10/8.00	0.397
	SY	89.65	3.60/6.75	0.311	89.60	3.50/6.90	0.399
5,000	EP	90.15	4.35/5.50	0.145	89.25	4.20/6.55	0.178
	RT	89.75	3.90/6.35	0.145	88.90	3.95/7.15	0.178
	SY	90.15	4.10/5.75	0.145	89.00	4.10/6.90	0.179
10,000	EP	89.60	4.70/5.70	0.103	89.45	4.95/5.60	0.126
	RT	89.85	4.20/5.95	0.103	89.70	4.45/5.85	0.126
	SY	89.95	4.30/5.75	0.103	89.40	4.90/5.70	0.126

Table 5.2: The table reports distinct features of the fixed-design bootstrap confidence intervals for the conditional ES at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes, the interval's average coverage rates (in %), the average rate of the conditional ES being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. The DGP is a GARCH(1,1) with **Gaussian innovations**.

highest sample sizes. An explanation for this result can be found in Gao and Song (2008, Remark 3.3) who assert that the effective sample size for the estimation of ES is solely $n\alpha$. All in all, we conclude that larger sample sizes are needed to obtain acceptable coverage probabilities.

Table 5.4 considers an increase in the interval's nominal value from 90% to 95%. Once again we conclude that the results are qualitatively similar to the benchmark. The average lengths of the intervals are larger for every sample size considered. These results are to be expected for bootstrap intervals with higher nominal value.

It might be of interest to compare the results for the conditional ES with those reported in Chapter 4 for the conditional VaR. There we find that the EP interval performs worse than the RT interval in small samples, which is in line with the theoretical findings in Falk and Kaufmann (1991). This result does not carry over to the conditional ES, since in most instances (except Table 5.4) the EP interval even outperforms the RT interval. To make a full comparison, we also computed

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
		low persistence			high persistence		
500	EP	73.75	4.75/21.50	1.183	72.25	5.45/22.30	1.566
	RT	71.30	0.75/27.95	1.183	70.75	1.00/28.25	1.566
	SY	74.85	1.55/23.60	1.226	73.70	2.00/24.30	1.616
1,000	EP	81.90	2.55/15.55	0.944	81.55	2.80/15.65	1.260
	RT	80.55	0.85/18.60	0.944	79.45	1.30/19.25	1.260
	SY	82.20	1.25/16.55	0.955	81.20	1.55/17.25	1.270
5,000	EP	87.05	3.15/9.80	0.476	88.30	2.30/9.40	0.632
	RT	87.35	2.55/10.10	0.476	88.00	2.50/9.50	0.632
	SY	87.30	2.85/9.85	0.477	88.25	2.40/9.35	0.632
10,000	EP	88.45	3.70/7.85	0.347	89.10	3.00/7.90	0.458
	RT	88.40	3.75/7.85	0.347	88.45	3.40/8.15	0.458
	SY	88.65	3.60/7.75	0.347	88.55	3.20/8.25	0.458

Table 5.3: The table reports distinct features of the fixed-design bootstrap confidence intervals for the conditional ES at **level $\alpha = 0.01$** with **nominal coverage $1 - \gamma = 90\%$** . For each interval type and different sample sizes, the interval’s average coverage rates (in %), the average rate of the conditional ES being below/above the interval (in %) and the interval’s average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. The DGP is a GARCH(1,1) with (normalized) **Student- t innovations (6 degrees of freedom)**.

the results where the DGP is a T-GARCH(1, 1). Results are vastly comparable and available upon request.

To summarize, the simulation study suggests that the fixed-design bootstrap works well in terms of average coverage. In comparison to the conditional VaR, higher sample sizes are necessary to obtain coverage rates close to the nominal value. There is no clear evidence to have a preference for any of the three intervals based on the simulation results in all different settings. This directly contrasts results for the conditional VaR in Chapter 4 for which the RT bootstrap interval is found superior.

5.6 Conclusion

We study the two-step estimation procedure of Francq and Zakoïan (2015) associated with the conditional ES. In the first step, the conditional volatility parameters are estimated by QMLE, while the second step corresponds to the estimation of

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
		low persistence			high persistence		
500	EP	89.40	1.85/8.75	0.732	89.65	2.10/8.25	0.984
	RT	90.65	1.00/8.35	0.732	90.85	1.00/8.15	0.984
	SY	90.50	1.15/8.35	0.745	91.10	1.30/7.60	0.997
1,000	EP	92.80	1.45/5.75	0.534	93.05	1.15/5.80	0.727
	RT	93.50	1.15/5.35	0.534	93.35	1.20/5.45	0.727
	SY	93.60	1.00/5.40	0.540	93.40	1.15/5.45	0.730
5,000	EP	93.80	2.25/3.95	0.257	94.35	1.60/4.05	0.342
	RT	94.45	2.05/3.50	0.257	94.95	1.50/3.55	0.342
	SY	94.20	2.10/3.70	0.257	94.60	1.55/3.85	0.342
10,000	EP	94.35	2.45/3.20	0.185	94.25	1.85/3.90	0.243
	RT	95.25	2.35/2.40	0.185	94.50	2.05/3.45	0.243
	SY	94.90	2.50/2.60	0.185	94.70	1.85/3.45	0.243

Table 5.4: The table reports distinct features of the fixed-design bootstrap confidence intervals for the conditional ES at **level $\alpha = 0.05$** with **nominal coverage $1 - \gamma = 95\%$** . For each interval type and different sample sizes, the interval's average coverage rates (in %), the average rate of the conditional ES being below/above the interval (in %) and the interval's average length are tabulated. The intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. The DGP is a GARCH(1,1) with (normalized) **Student- t innovations (6 degrees of freedom)**.

conditional ES based on the first-step residuals. We find that the estimators of the parameters that comprise the conditional ES have a joint asymptotic distribution that does not depend on the innovation density. This is in direct contrast with the conditional VaR estimator for which density estimation is required.

A fixed-design residual bootstrap method is proposed to mimic the finite sample distribution of the two-step estimator and its consistency is proven under mild assumptions. In addition, an algorithm is provided for the construction of bootstrap intervals for the conditional ES to take into account the uncertainty induced by estimation. Three interval types are suggested and a simulation study is conducted to investigate their performance in finite samples. Firstly, we find that average coverage rates of all intervals are close to nominal value, except when sample size is low. Secondly, we find that there is no clear evidence that any of the proposed intervals outperforms the others. This contrasts the results in Chapter 4 for the conditional VaR, where we find superiority of the reversed-tails bootstrap interval.

5.A Auxiliary Results and Proofs

Lemma 5.1. *Suppose Assumptions 5.1–5.10 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have $A_n^* \xrightarrow{P^*} 0$ in probability.*

Proof. The proof is inspired by Chen (2008, proof of Lemma 2). Take $\delta \in (1/4, 1/2)$, and expand

$$\begin{aligned}
 A_n^* &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}) (\mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) \\
 &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}) \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \\
 &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}) \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \\
 &\quad \underbrace{\times \mathbb{1}_{\{|\hat{\eta}_t^* - \eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| < n^{-\delta}\}}}_I \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}) \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} - \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \\
 &\quad \underbrace{\times \mathbb{1}_{\{|\hat{\eta}_t^* - \eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \geq n^{-\delta}\}}}_{II}.
 \end{aligned}$$

The first term can be bounded by

$$\begin{aligned}
 |I| &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}| \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} + \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \\
 &\quad \times \mathbb{1}_{\{|\hat{\eta}_t^* - \eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| < n^{-\delta}\}} \\
 &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n (n^{-\delta} + |\eta_t^* - \hat{\xi}_{n,\alpha}|) \underbrace{\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\eta_t^* - n^{-\delta} < \hat{\xi}_{n,\alpha} + n^{-\delta}\}}}_{=\mathbb{1}_{\{0 \leq \eta_t^* - \hat{\xi}_{n,\alpha} < 2n^{-\delta}\}}} \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n (n^{-\delta} + |\eta_t^* - \hat{\xi}_{n,\alpha}|) \underbrace{\mathbb{1}_{\{\hat{\xi}_{n,\alpha} - n^{-\delta} \leq \eta_t^* + n^{-\delta}\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}}_{=\mathbb{1}_{\{-2n^{-\delta} \leq \eta_t^* - \hat{\xi}_{n,\alpha} < 0\}}} \\
 &\leq \frac{2}{\sqrt{n}} \sum_{t=1}^n (n^{-\delta} + |\eta_t^* - \hat{\xi}_{n,\alpha}|) \mathbb{1}_{\{|\eta_t^* - \hat{\xi}_{n,\alpha}| \leq 2n^{-\delta}\}}
 \end{aligned}$$

$$\leq \underbrace{\frac{6}{\sqrt{nn}^\delta} \sum_{t=1}^n \mathbb{1}_{\{|\eta_t^* - \hat{\xi}_{n,\alpha}| \leq 2n^{-\delta}\}}}_{W_n^*}$$

Note that, given the original sample, the random variables $X_{t,n}^* = \mathbb{1}_{\{|\eta_t^* - \hat{\xi}_{n,\alpha}| < 2n^{-\delta}\}}$ are i.i.d. The conditional mean satisfies $\mathbb{E}^*[X_{t,n}^*] = O_p(n^{-\delta})$; to appreciate why, we have

$$\begin{aligned} \mathbb{E}^*[X_{t,n}^*] &= \mathbb{P}^*[\hat{\xi}_{n,\alpha} - 2n^{-\delta} < \eta_t^* < \hat{\xi}_{n,\alpha} + 2n^{-\delta}] \\ &= \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} + 2n^{-\delta}) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} - 2n^{-\delta}) \\ &= F(\hat{\xi}_{n,\alpha} + 2n^{-\delta}) - F(\hat{\xi}_{n,\alpha} - 2n^{-\delta}) + o_p(n^{-1/2}) \\ &= 4n^{-\delta} f(\hat{\xi}_{n,\alpha} + b_n) + o_p(n^{-1/2}) = 4n^{-\delta} f(\xi_\alpha) + o_p(n^{-\delta}) + o_p(n^{-1/2}) \end{aligned}$$

where $b_n \in (-2n^{-\delta}, 2n^{-\delta})$. The third equality follows from Lemma 4.3. The mean value theorem is applied to obtain the fourth equality and the last equality is due to continuity of f in a neighborhood of ξ_α and $\hat{\xi}_{n,\alpha} + b_n \xrightarrow{a.s.} \xi_\alpha$. Thus, we have

$$\begin{aligned} \mathbb{E}^*[W_n^*] &= \frac{6}{\sqrt{nn}^\delta} \sum_{t=1}^n \mathbb{E}^*[X_{t,n}^*] = O_p(\sqrt{n}/n^{2\delta}) = O_p(n^{-\delta}) \\ \text{Var}^*[W_n^*] &= \frac{36}{n^{1+2\delta}} \sum_{t=1}^n \mathbb{E}^*[X_{t,n}^*] \underbrace{\left(1 - \mathbb{E}^*[X_{t,n}^*]\right)}_{\leq 1} = O_p(n^{-3\delta}) \end{aligned}$$

implying $W_n^* \xrightarrow{p^*} 0$ in probability and we conclude that $I \xrightarrow{p^*} 0$ in probability. Regarding the second term, we write $\hat{\mathbb{I}}_{n,t}^* = \mathbb{1}_{\{|\hat{\eta}_t^* - \eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \geq n^{-\delta}\}}$ and establish the following bound

$$\begin{aligned} |II| &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}| \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} + \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \hat{\mathbb{I}}_{n,t}^* \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}| \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}\}} \right. \\ &\quad \left. + \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \hat{\eta}_t^*\}} \right) \hat{\mathbb{I}}_{n,t}^* \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{\eta}_t^* - \hat{\xi}_{n,\alpha}| \left(\mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \eta_t^*\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}^*\}} \mathbb{1}_{\{\hat{\xi}_{n,\alpha} \leq \hat{\eta}_t^*\}} \right. \\ &\quad \left. + \mathbb{1}_{\{\hat{\xi}_{n,\alpha}^* \leq \hat{\eta}_t^*\}} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \mathbb{1}_{\{\hat{\eta}_t^* < \hat{\xi}_{n,\alpha}\}} \right) \hat{\mathbb{I}}_{n,t}^* \end{aligned}$$

$$\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n 2|\hat{\eta}_t^* - \eta_t^*| \hat{\mathbb{1}}_{n,t}^* + \frac{1}{\sqrt{n}} \sum_{t=1}^n 2|\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \hat{\mathbb{1}}_{n,t}^*.$$

The Taylor expansion in (4.93) and $\hat{\eta}_t^* = \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} \eta_t^*$ gives

$$\begin{aligned} \eta_t^* - \hat{\eta}_t^* &= \hat{D}_t'(\hat{\theta}_n^* - \hat{\theta}_n) \eta_t^* \\ &\quad + \frac{1}{2}(\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \left(\tilde{H}_t(\hat{\theta}_n) - 2\tilde{D}_t(\hat{\theta}_n) \tilde{D}_t'(\hat{\theta}_n) \right) (\hat{\theta}_n^* - \hat{\theta}_n) \eta_t^* \end{aligned} \quad (5.21)$$

with $\check{\theta}_n$ between $\hat{\theta}_n^*$ and $\hat{\theta}_n$. Employing the Cauchy-Schwarz inequality we obtain

$$|\hat{\eta}_t^* - \eta_t^*| = |\hat{\theta}_n^* - \hat{\theta}_n| \left(\|\hat{D}_t\| |\eta_t^*| + \frac{1}{2} \|\hat{\theta}_n^* - \hat{\theta}_n\|^2 \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \left(\|\tilde{H}_t(\hat{\theta}_n)\| + 2\|\tilde{D}_t(\hat{\theta}_n)\|^2 \right) |\eta_t^*| \right).$$

Hence, $|II|$ can further be bounded by

$$\begin{aligned} |II| &\leq n \|\hat{\theta}_n^* - \hat{\theta}_n\|^2 \underbrace{\frac{1}{\sqrt{nn}} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n)} \left(\|\tilde{H}_t(\hat{\theta}_n)\| + 2\|\tilde{D}_t(\hat{\theta}_n)\|^2 \right) |\eta_t^*|}_{II_1} \\ &\quad + 2\sqrt{n} |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \underbrace{\frac{1}{n} \sum_{t=1}^n \hat{\mathbb{1}}_{n,t}^*}_{II_2} + 2\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| \underbrace{\frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\| |\eta_t^*| \hat{\mathbb{1}}_{n,t}^*}_{II_3}. \end{aligned}$$

Since $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} N(0, \frac{\kappa-1}{4} J^{-1})$ almost surely and $\sqrt{n}(\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}) \xrightarrow{d^*} N(0, \zeta_\alpha)$ in probability for some ζ_α (see Theorem 4.3), it remains to show that the under-braced terms converge in conditional probability to zero in probability. Consider II_1 ; for every $\varepsilon > 0$ we obtain

$$\begin{aligned} \mathbb{P}^*[II_1 \geq \varepsilon] &\leq \mathbb{P}^*[II_1 \geq \varepsilon \cap \check{\theta}_n \in \mathcal{V}(\theta_0)] + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ &\leq \mathbb{P}^* \left[\underbrace{\frac{1}{\sqrt{nn}} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left(\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) |\eta_t^*|}_{Q_{t,n}} \geq \varepsilon \right] \\ &\quad + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ &\leq \frac{1}{\varepsilon} \mathbb{E}^* \left[\frac{1}{\sqrt{nn}} \sum_{t=1}^n Q_{t,n} |\eta_t^*| \right] + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \end{aligned} \quad (5.22)$$

$$= \frac{\mathbb{E}^*[\eta_t^*]}{\varepsilon\sqrt{nn}} \sum_{t=1}^n Q_{t,n} + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)],$$

where the last inequality is due to Markov. As $\mathbb{E}^*[\eta_t^*] \leq (\mathbb{E}^*[\eta_t^{*2}])^{1/2} \xrightarrow{a.s.} (\mathbb{E}[\eta_t^2])^{1/2} = 1$, $\frac{1}{n} \sum_{t=1}^n Q_{t,n} = O_p(1)$ and $\mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \xrightarrow{a.s.} 0$ by Lemma 4.4 and the proof of Lemma 4.10, we have $\mathbb{P}^*[II_1 \geq \varepsilon] \xrightarrow{P^*} 0$ and we conclude that $II_1 \xrightarrow{P^*} 0$ in probability. Next, we focus on II_2 . For every $\varepsilon > 0$ we obtain

$$\begin{aligned} & \mathbb{P}^*[II_2 \geq \varepsilon] \\ & \leq \mathbb{P}^*\left[\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\hat{\eta}_t^* - \eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \geq n^{-\delta}\}} \geq \varepsilon \cap \check{\theta}_n \in \mathcal{V}(\theta_0)\right] \\ & \quad + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \mathbb{P}^*\left[\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{|\hat{\theta}_n^* - \hat{\theta}_n| \|\hat{D}_t\| |\eta_t^*| + \frac{1}{2} \|\hat{\theta}_n^* - \hat{\theta}_n\|^2 Q_{n,t} |\eta_t^*| + |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| \geq n^{-\delta}\}} \geq \varepsilon\right] \\ & \quad + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \mathbb{P}^*\left[\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\frac{C}{\sqrt{n}} \|\hat{D}_t\| |\eta_t^*| + \frac{C^2}{2n} Q_{n,t} |\eta_t^*| + \frac{C}{\sqrt{n}} \geq n^{-\delta}\}} \geq \varepsilon\right] \\ & \quad + \mathbb{P}^*[\sqrt{n} |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| > C] + \mathbb{P}^*[\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > C] + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \frac{1}{\varepsilon n} \sum_{t=1}^n \mathbb{P}^*\left[\frac{C}{\sqrt{n}} \|\hat{D}_t\| |\eta_t^*| + \frac{C^2}{2n} Q_{n,t} |\eta_t^*| + \frac{C}{\sqrt{n}} \geq n^{-\delta}\right] \\ & \quad + \mathbb{P}^*[\sqrt{n} |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| > C] + \mathbb{P}^*[\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > C] + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & \leq \frac{n^\delta}{\varepsilon n} \sum_{t=1}^n \mathbb{E}^*\left[\frac{C}{\sqrt{n}} \|\hat{D}_t\| |\eta_t^*| + \frac{C^2}{2n} Q_{n,t} |\eta_t^*| + \frac{C}{\sqrt{n}}\right] \\ & \quad + \mathbb{P}^*[\sqrt{n} |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| > C] + \mathbb{P}^*[\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > C] + \mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\ & = \underbrace{\frac{n^\delta}{\sqrt{n}} \frac{C \mathbb{E}^*[\eta_t^*]}{\varepsilon} \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|}_{II_{2,1}} + \underbrace{\frac{n^\delta}{n} \frac{C^2 \mathbb{E}^*[\eta_t^*]}{2\varepsilon} \frac{1}{n} \sum_{t=1}^n Q_{n,t}}_{II_{2,2}} + \underbrace{\frac{n^\delta}{\sqrt{n}} \frac{C}{\varepsilon}}_{II_{2,3}} \\ & \quad + \underbrace{\mathbb{P}^*[\sqrt{n} |\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}| > C]}_{II_{2,4}} + \underbrace{\mathbb{P}^*[\sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > C]}_{II_{2,5}} + \underbrace{\mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)]}_{II_{2,6}}. \end{aligned}$$

Previously, we have shown that $II_{2,6} \xrightarrow{a.s.} 0$, whereas the $II_{2,4}$ and $II_{2,5}$ can be made arbitrarily small in probability by choosing C sufficiently large. Given C ,

we find $II_{2,3} \rightarrow 0$ as $\frac{n^\delta}{\sqrt{n}} \rightarrow 0$. Recalling that $\lim_{n \rightarrow \infty} \mathbb{E}^* [|\eta_t^*|] \leq 1$ almost surely and $\sum_{t=1}^n Q_{n,t} = O_p(1)$ we find $II_{2,2} \xrightarrow{P} 0$. The Cauchy-Schwarz inequality gives $\frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\| \leq \left(\frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 \right)^{1/4}$ and together with $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 < \infty$ almost surely (see proof of Lemma 4.7), we get that $II_{2,1} \xrightarrow{a.s.} 0$. Thus, $\mathbb{P}^* [II_2 \geq \varepsilon] \xrightarrow{P} 0$ and we establish $II_2 \xrightarrow{P^*} 0$ in probability. Regarding II_3 , Hölder's inequality implies

$$II_3 \leq \left(\frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 \right)^{\frac{1}{4}} \underbrace{\left(\frac{1}{n} \sum_{t=1}^n \hat{\eta}_{n,t}^* \right)^{\frac{1}{4}}}_{=II_2}.$$

We have $\frac{1}{n} \sum_{t=1}^n |\eta_t^*|^2 \xrightarrow{P^*} \mathbb{E}[\eta_t^2]$ almost surely as

$$\mathbb{E}^* \left[\frac{1}{n} \sum_{t=1}^n |\eta_t^*|^2 \right] = \mathbb{E}^* [|\eta_t^*|^2] \xrightarrow{a.s.} \mathbb{E}[\eta_t^2]$$

and

$$\mathbb{V}\text{ar}^* \left[\frac{1}{n} \sum_{t=1}^n |\eta_t^*|^2 \right] = \frac{1}{n} \mathbb{V}\text{ar}^* [|\eta_t^*|^2] \xrightarrow{a.s.} 0$$

by Lemma 4.4. Recalling that $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 < \infty$ almost surely and $II_2 \xrightarrow{P^*} 0$ in probability we establish $II_3 \xrightarrow{P^*} 0$ in probability. \square

Lemma 5.2. *Suppose Assumptions 5.1–5.4, 5.5(i), 5.5(iii), 5.6, 5.7, 5.9 and 5.10 hold with $a = \pm 12$, $b = 12$ and $c = 6$. Then, we have $C_n^* = \alpha \mu_\alpha \Omega \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{P^*}(1)$ in probability.*

Proof. Inserting (5.21) into the definition of C_n^* we obtain

$$\begin{aligned} C_n^* &= \underbrace{-\frac{1}{n} \sum_{t=1}^n \hat{D}_t' \eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)}_I \\ &\quad - \underbrace{\frac{1}{2} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{1}{\sqrt{nn}} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left(\tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n) \tilde{D}_t'(\check{\theta}_n) \right) \eta_t^*}_{II} \\ &\quad \times \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n). \end{aligned}$$

Since $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} N(0, \frac{\kappa-1}{4}J^{-1})$ almost surely by Proposition 4.1 and $II \xrightarrow{P^*} 0$ in probability by (5.22), it remains to show that $I \xrightarrow{P^*} \alpha\mu_\alpha\Omega'$ in probability. Noting that $-\mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] = \alpha\mu_\alpha$, we obtain

$$\begin{aligned}\mathbb{E}^*[I] &= -\mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \frac{1}{n} \sum_{t=1}^n \hat{D}'_t = -\mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}'_n \xrightarrow{a.s.} \alpha\mu_\alpha\Omega' \\ \mathbb{V}\text{ar}^*[I] &= \mathbb{V}\text{ar}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \frac{1}{n^2} \sum_{t=1}^n \hat{D}_t \hat{D}'_t = \frac{1}{n} \mathbb{V}\text{ar}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{J}_n \xrightarrow{a.s.} 0\end{aligned}$$

by Lemmas 4.2 and 4.4, which completes the proof. \square

Lemma 5.3. *Suppose Assumptions 5.1–5.4, 5.5(i), 5.5(iii), 5.6, 5.9 and 5.10 hold with $a = -1, 4$, $b = 4$ and $c = 2$. Then, we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{\hat{D}_t(\eta_t^{*2} - 1)}{(\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} + \alpha_n(\hat{\xi}_{n,\alpha} + \hat{\mu}_{n,\alpha})} \right) \xrightarrow{d^*} N(0, \Psi_\alpha)$$

almost surely with

$$\Psi_\alpha = \begin{pmatrix} (\kappa - 1)J & \alpha x_\alpha \Omega \\ \alpha x_\alpha \Omega' & \alpha^2 \sigma_\alpha^2 \end{pmatrix}.$$

Proof. In the proof of Lemma 4.7 it is shown that $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\mathbb{E}^*[\eta_t^{*2}] - 1) = 0$ for sufficiently large n almost surely since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0 \in \mathring{\Theta}$ and $\mathbb{E}^*[\eta_t^{*2}] = 1$ whenever $\hat{\theta}_n \in \mathring{\Theta}$ under Assumption 5.10. It remains to show that for each $\lambda = (\lambda'_1, \lambda'_2)' \in \mathbb{R}^{r+1}$ with $\|\lambda\| \neq 0$

$$\sum_{t=1}^n \frac{1}{\sqrt{n}} \lambda' \underbrace{\left(\frac{\hat{D}_t(\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}])}{(\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} + \alpha_n(\hat{\xi}_{n,\alpha} + \hat{\mu}_{n,\alpha})} \right)}_{Z_{n,t}^*} \xrightarrow{d^*} N(0, \lambda' \Psi_\alpha \lambda)$$

almost surely by the Cramér-Wold device. By construction, we have $\mathbb{E}^*[Z_{n,t}^*] = 0$. Further, we have that $s_n^2 = \sum_{t=1}^n \mathbb{V}\text{ar}^*[Z_{n,t}^*]$ is equal to

$$\lambda' \begin{pmatrix} \mathbb{V}\text{ar}^*[\eta_t^{*2}] \hat{J}_n & \text{Cov}^*[\eta_t^{*2}, (\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}'_n \\ \text{Cov}^*[\eta_t^{*2}, (\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}'_n & \mathbb{V}\text{ar}^*[(\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{J}_n \end{pmatrix} \lambda.$$

Lemma 4.2 gives $\hat{J}_n \xrightarrow{a.s.} J$ and $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$. Further, 4.5 implies

$$\begin{aligned}
 & \mathbb{V}\text{ar}^*[(\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \\
 &= \mathbb{E}^*[\eta_t^{*2} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \left(\mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \right)^2 \\
 & \quad + \hat{\xi}_{n,\alpha}^2 \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \left(1 - \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \right) \\
 & \quad - 2\hat{\xi}_{n,\alpha} \left(\mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \right) \\
 & \xrightarrow{a.s.} \mathbb{E}[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \left(\mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \right)^2 + \xi_\alpha^2 \mathbb{E}[\mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \left(1 - \mathbb{E}[\mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \right) \\
 & \quad - 2\xi_\alpha \left(\mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \mathbb{E}[\mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \right) \\
 &= \mathbb{V}\text{ar}[(\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]
 \end{aligned}$$

and

$$\begin{aligned}
 & \mathbb{C}\text{ov}^*[\eta_t^{*2}, (\eta_t^* - \hat{\xi}_{n,\alpha}) \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \\
 &= \mathbb{E}^*[\eta_t^{*3} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \mathbb{E}^*[\eta_t^{*2}] \mathbb{E}^*[\eta_t^* \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \\
 & \quad - \hat{\xi}_{n,\alpha} \left(\mathbb{E}^*[\eta_t^{*2} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \mathbb{E}^*[\eta_t^{*2}] \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \right) \\
 & \xrightarrow{a.s.} \mathbb{E}[\eta_t^3 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \mathbb{E}[\eta_t^2] \mathbb{E}[\eta_t \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \\
 & \quad - \xi_\alpha \left(\mathbb{E}[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \mathbb{E}[\eta_t^2] \mathbb{E}[\mathbb{1}_{\{\eta_t < \xi_\alpha\}}] \right) \\
 &= \mathbb{C}\text{ov}[\eta_t^2, (\eta_t - \xi_\alpha) \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]
 \end{aligned}$$

as well as

$$\mathbb{V}\text{ar}^*[\eta_t^{*2}] = \mathbb{E}^*[\eta_t^{*4}] - \left(\mathbb{E}[\eta_t^{*2}] \right)^2 \xrightarrow{a.s.} \kappa - 1.$$

Thus, we get $s_n^2 \xrightarrow{a.s.} \lambda' \Psi_\alpha \lambda$. Next, we verify Lindeberg condition. For any $\varepsilon > 0$

$$\begin{aligned}
 & \sum_{t=1}^n \mathbb{E}^* [Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}}] \\
 & \leq \underbrace{\sum_{t=1}^n \mathbb{E}^* [Z_{n,t}^{*2} \mathbb{1}_{\{|\eta_t^*| > C\}}]}_I + \underbrace{\sum_{t=1}^n \mathbb{E}^* [Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}} \mathbb{1}_{\{|\eta_t^*| \leq C\}}]}_{II}
 \end{aligned}$$

holds, where $C > 0$. Employing the elementary inequalities $(x + y)^z \leq 2^z(x^z + y^z)$ and $|x - y|^z \leq x^z + y^z$ for all $x, y, z \geq 0$ we find that

$$Z_{n,t}^{*2} \leq \frac{8}{n} \left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 (\eta_t^{*2} + \hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2) \right).$$

Thus, we obtain

$$\begin{aligned} I &\leq \frac{8}{n} \sum_{t=1}^n \mathbb{E}^* \left[\left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 (\eta_t^{*2} + \hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2) \right) \mathbb{1}_{\{|\eta_t^*| > C\}} \right] \\ &= 8 \left(\lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^* [\eta_t^{*4} \mathbb{1}_{\{|\eta_t^*| > C\}}] + \lambda_2^2 E^* [\eta_t^{*2} \mathbb{1}_{\{|\eta_t^*| > C\}}] \right. \\ &\quad \left. + (\lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^* [\eta_t^{*2}]^2 + \lambda_2^2 (\hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2)) \mathbb{E}^* [\mathbb{1}_{\{|\eta_t^*| > C\}}] \right) \\ &\stackrel{a.s.}{\rightarrow} 8 \left(\lambda'_1 J \lambda_1 \mathbb{E} [\eta_t^4 \mathbb{1}_{\{|\eta_t| > C\}}] + \lambda_2^2 E [\eta_t^2 \mathbb{1}_{\{|\eta_t| > C\}}] \right. \\ &\quad \left. + (\lambda'_1 J \lambda_1 \mathbb{E} [\eta_t^2]^2 + \lambda_2^2 (\xi_\alpha^2 + \mu_\alpha^2)) \mathbb{E} [\mathbb{1}_{\{|\eta_t| > C\}}] \right) \end{aligned}$$

and choosing C sufficiently large yields $I \stackrel{a.s.}{\rightarrow} 0$. Given a value of C , we have

$$\begin{aligned} II &\leq \frac{8}{n} \sum_{t=1}^n \mathbb{E}^* \left[\left((\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 (\eta_t^{*2} + \hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2) \right) \right. \\ &\quad \left. \times \mathbb{1}_{\{||\lambda_1| |(\eta_t^{*2} + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t| | + |\lambda_2| (|\eta_t^*| + |\hat{\xi}_{n,\alpha}| + |\hat{\mu}_{n,\alpha}|) \geq \sqrt{n} s_n \epsilon\}} \mathbb{1}_{\{|\eta_t^*| \leq C\}} \right] \\ &\leq \frac{8}{n} \sum_{t=1}^n \left((\lambda'_1 \hat{D}_t)^2 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 (C^2 + \hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2) \right) \\ &\quad \times \mathbb{1}_{\{||\lambda_1| |(C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t| | + |\lambda_2| (C + |\hat{\xi}_{n,\alpha}| + |\hat{\mu}_{n,\alpha}|) \geq \sqrt{n} s_n \epsilon\}} \\ &\leq 8 \left(\lambda'_1 \hat{J}_n \lambda_1 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 (C^2 + \hat{\xi}_{n,\alpha}^2 + \hat{\mu}_{n,\alpha}^2) \right) \\ &\quad \times \mathbb{1}_{\{||\lambda_1| |(C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t ||\hat{D}_t| | + |\lambda_2| (C + |\hat{\xi}_{n,\alpha}| + |\hat{\mu}_{n,\alpha}|) \geq \sqrt{n} s_n \epsilon\}} \\ &\stackrel{a.s.}{\rightarrow} 8 \left(\lambda'_1 J \lambda_1 (C^4 + \mathbb{E}[\eta_t^2]^2) + \lambda_2^2 (C^2 + \xi_\alpha^2 + \mu_\alpha^2) \right) \times 0 = 0 \end{aligned}$$

as $\max_t ||\hat{D}_t| | / \sqrt{n} \stackrel{a.s.}{\rightarrow} 0$. Combining results, gives $\frac{1}{s_n^2} \sum_{t=1}^n \mathbb{E}^* [Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \epsilon\}}] \stackrel{a.s.}{\rightarrow} 0$. The Central Limit Theorem for triangular arrays (cf. Billingsley, 1986, Theorem 27.3) implies that $\sum_{t=1}^n Z_{n,t}^*$ converges in conditional distribution to $N(0, \lambda' \Psi_\alpha \lambda)$ almost surely, which completes the proof. \square

5.B Derivation of Analytical Expressions

Let $Y \sim t_\nu$ with cdf F_ν and pdf f_ν , where t_ν denotes the Student- t distribution with ν degrees of freedom. Recall that $\sigma_\nu^2 = \frac{\nu-2}{\nu}$ such that $\eta = \sigma_\nu Y$ is now appropriately standardized, i.e. $\mathbb{E}[\eta^2] = 1$. Then we get

$$\begin{aligned} F(x) &= \mathbb{P}[\eta \leq x] = \mathbb{P}[Y \leq x/\sigma_\nu] = F_\nu(x/\sigma_\nu) \\ \xi_\alpha &= F^{-1}(\alpha) = \sigma_\nu F_\nu^{-1}(\alpha). \end{aligned}$$

The following relationship links the (conditional) moments of η and Y :

$$\mathbb{E}[\eta^m | \eta < \xi_\alpha] = \sigma_\nu^m \mathbb{E}[Y^m | Y < F_\nu^{-1}(\alpha)] \quad (5.23)$$

with $m \in \mathbb{N}$. Using moments of the truncated Student- t distribution derived in Kim (2008, p. 84) we can find closed form expressions for the conditional expectations of Y^m . For $m = 1$ and any $b \in \mathbb{R}$ we have⁶

$$\begin{aligned} \mathbb{E}[Y | Y < b] &= -\frac{\Gamma(\frac{\nu-1}{2})\nu^{\nu/2}}{2F_\nu(b)\Gamma(\frac{\nu}{2})\sqrt{\pi}}(\nu + b^2)^{-\frac{\nu-1}{2}} \\ &= -\frac{\Gamma(\frac{\nu-1}{2})\sqrt{\nu}}{2F_\nu(b)\frac{\nu-2}{2}\Gamma(\frac{\nu-2}{2})\sqrt{\pi}}\left(1 + \frac{b^2}{\nu}\right)^{-\frac{\nu-1}{2}} \\ &= -\frac{\Gamma(\frac{\nu-1}{2})}{F_\nu(b)\sigma_\nu\sqrt{\nu-2}\Gamma(\frac{\nu-2}{2})\sqrt{\pi}}\left(1 + \frac{(\sigma_\nu b)^2}{\nu-2}\right)^{-\frac{\nu-1}{2}} \\ &= -\frac{f_{\nu-2}(\sigma_\nu b)}{F_\nu(b)\sigma_\nu}, \end{aligned}$$

where we recognize that $\Gamma(\frac{\nu}{2}) = \frac{\nu-2}{2}\Gamma(\frac{\nu-2}{2})$. Together with (5.23) we have

$$\mathbb{E}[\eta | \eta < \xi_\alpha] = \sigma_\nu \mathbb{E}[Y | Y < F_\nu^{-1}(\alpha)] = -\frac{f_{\nu-2}(\xi_\alpha)}{\alpha}. \quad (5.24)$$

Similarly, we can derive for $m = 2$ and any $b \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}[Y^2 | Y < b] &= \frac{1}{\sigma_\nu^2} - b \frac{\Gamma(\frac{\nu-1}{2})\nu^{\nu/2}}{2F_\nu(b)\Gamma(\frac{\nu}{2})\sqrt{\pi}}(\nu + b^2)^{-\frac{\nu-1}{2}} \\ &= \frac{1}{\sigma_\nu^2} + b \mathbb{E}[Y | Y < b]. \end{aligned}$$

⁶We only truncate from above, hence the lower truncation bound of Kim (2008) is $a = -\infty$.

Combined with (5.23) we arrive at

$$\begin{aligned}
 \mathbb{E}[\eta^2 | \eta < \xi_\alpha] &= \sigma_\nu^2 \mathbb{E}[Y^2 | Y < F_\nu^{-1}(\alpha)] \\
 &= \sigma_\nu^2 \left(\frac{1}{\sigma_\nu^2} + F_\nu^{-1}(\alpha) \mathbb{E}[Y | Y < F_\nu^{-1}(\alpha)] \right) \\
 &= 1 + \sigma_\nu^2 F_\nu^{-1}(\alpha) \mathbb{E}[Y | Y < F_\nu^{-1}(\alpha)] \\
 &= 1 - \xi_\alpha \frac{f_{\nu-2}(\xi_\alpha)}{\alpha}.
 \end{aligned}$$

Now, as $\mathbb{P}[\eta < \xi_\alpha] = \alpha$, we obtain

$$\mathbb{E}[\eta^2 \mathbb{1}_{\{\eta < \xi_\alpha\}}] - \alpha = \mathbb{E}[\eta^2 | \eta < \xi_\alpha] \mathbb{P}[\eta < \xi_\alpha] - \alpha = -\xi_\alpha f_{\nu-2}(\xi_\alpha). \quad (5.25)$$

Finally, we consider $m = 3$. Using $\Gamma(\frac{\nu-1}{2}) = \frac{\nu-3}{2} \Gamma(\frac{\nu-3}{2})$ or equivalently $\Gamma(\frac{\nu-3}{2}) = \frac{2}{\nu-3} \Gamma(\frac{\nu-1}{2})$, it follows that

$$\begin{aligned}
 \mathbb{E}[Y^3 | Y < b] &= -\frac{\Gamma(\frac{\nu-3}{2}) \nu^{\nu/2}}{2F_\nu(b) \Gamma(\frac{\nu}{2}) \sqrt{\pi}} (\nu + b^2)^{-\frac{\nu-3}{2}} \\
 &\quad - b^2 \frac{\Gamma(\frac{\nu-1}{2}) \nu^{\nu/2}}{2F_\nu(b) \Gamma(\frac{\nu}{2}) \sqrt{\pi}} (\nu + b^2)^{-\frac{\nu-1}{2}} \\
 &= -\left(\frac{2(\nu + b^2)}{\nu - 3} + b^2 \right) \frac{\Gamma(\frac{\nu-1}{2}) \nu^{\nu/2}}{2F_\nu(b) \Gamma(\frac{\nu}{2}) \sqrt{\pi}} (\nu + b^2)^{-\frac{\nu-1}{2}} \\
 &= \left(\frac{2(\nu + b^2)}{\nu - 3} + b^2 \right) \mathbb{E}[Y | Y < b],
 \end{aligned}$$

which leads to

$$\begin{aligned}
 \mathbb{E}[\eta^3 | \eta < \xi_\alpha] &= \sigma_\nu^3 \mathbb{E}[Y^3 | Y < F_\nu^{-1}(\alpha)] \\
 &= \sigma_\nu^3 \left(\frac{2(\nu + \sigma_\nu^{-2} \xi_\alpha^2)}{\nu - 3} + \left(\frac{\xi_\alpha}{\sigma_\nu} \right)^2 \right) \mathbb{E}[Y | Y < F_\nu^{-1}(\alpha)] \\
 &= \left(\frac{2(\nu \sigma_\nu^2 + \xi_\alpha^2)}{\nu - 3} + \xi_\alpha^2 \right) \sigma_\nu \mathbb{E}[Y | Y < F_\nu^{-1}(\alpha)] \\
 &= -\left(\frac{2(\nu \sigma_\nu^2 + \xi_\alpha^2)}{\nu - 3} + \xi_\alpha^2 \right) \frac{f_{\nu-2}(\xi_\alpha)}{\alpha}.
 \end{aligned}$$

Thus, we obtain

$$\begin{aligned}\mathbb{E}[\eta^3 \mathbb{1}_{\{\eta < \xi_\alpha\}}] &= \mathbb{E}[\eta^3 | \eta < \xi_\alpha] \mathbb{P}[\eta < \xi_\alpha] \\ &= -\left(\frac{2(\nu\sigma_\nu^2 + \xi_\alpha^2)}{\nu - 3} + \xi_\alpha^2\right) f_{\nu-2}(\xi_\alpha).\end{aligned}\tag{5.26}$$

From (5.24) – (5.26), we get the following expressions for the quantities μ_α , p_α and q_α :

$$\begin{aligned}\mu_\alpha &= -\mathbb{E}[\eta | \eta < \xi_\alpha] = \frac{f_{\nu-2}(\xi_\alpha)}{\alpha} \\ p_\alpha &= \mathbb{E}[\eta^2 \mathbb{1}_{\{\eta < \xi_\alpha\}}] - \alpha = -\xi_\alpha f_{\nu-2}(\xi_\alpha) \\ q_\alpha &= \mathbb{E}[\eta^3 \mathbb{1}_{\{\eta < \xi_\alpha\}}] = -\left(\frac{2(\nu\sigma_\nu^2 + \xi_\alpha^2)}{\nu - 3} + \xi_\alpha^2\right) f_{\nu-2}(\xi_\alpha).\end{aligned}$$

Note that the Student- t distribution approaches the standard normal distribution as $\nu \rightarrow \infty$. In that case, $\sigma_\nu \rightarrow 1$ and also $f_\nu(\cdot) \rightarrow \phi(\cdot)$ and $F_\nu(\cdot) \rightarrow \Phi(\cdot)$, i.e. the standard normal pdf and cdf, respectively. Hence, when η is standard normally distributed, we have $\xi_\alpha = \Phi^{-1}(\alpha)$ as well as $\mu_\alpha = \frac{\phi(\xi_\alpha)}{\alpha}$, $p_\alpha = -\xi_\alpha \phi(\xi_\alpha)$ and $q_\alpha = -(2 + \xi_\alpha^2)\phi(\xi_\alpha)$.

Chapter 6

Conclusion

“To succeed, jump as quickly at opportunities as you do at conclusions.”

-Benjamin Franklin (1706-1790)

This chapter provides an overall conclusion of the thesis. More detailed, chapter-specific conclusions can be found at the end of each of the core chapters. Risk measures play a key role in financial risk management and are enforced by current legislation to protect financial stability. In particular VaR and ES are popular risk measures, that are used to calculate capital reserves and to assess the market risks associated with financial assets. Frequently these risk measures are applied conditionally to account for the temporal dependence of financial data and point estimates are obtained. To quantify the uncertainty around these point estimates, practitioners often construct confidence intervals by resorting to bootstrap methods. This thesis provides a theoretical justification for commonly constructed intervals around point estimates of conditional objects such as conditional risk measures. New bootstrap methods are proposed to quantify the uncertainty around the point estimates of the conditional VaR and ES and theoretical results are presented confirming their validity. Their performances are evaluated by means of simulations and practical recommendations are made to practitioners who choose to apply these bootstrap methods in empirical applications.

Chapter 2 focuses on quantifying the parameter uncertainty around point estimates of conditional objects. A fundamental issue arises stemming from the fact that on one hand one must condition on the sample as the past informs about the present, yet on the other hand one must allow the data to be treated as random to account for estimation uncertainty. Typically confidence intervals are constructed using an unrealistic assumption of observing two independent processes, which bypasses the issue. An alternative, realistic justification for this type of intervals is presented based on a sample-split approach and a weak dependence condition. To acknowledge that the conditional quantities vary over time, a merging concept is employed, which generalizes the notion of weak convergence.

Chapter 3 highlights the general setup of the previous chapter. It formally verifies the high level assumptions of Chapter 2 for various popular time series models. Among others these include the conditional mean in an AR(1) and ARMA(1,1) model, the conditional variance in a GARCH(1,1) model as well as the conditional volatility in a threshold GARCH model. In addition, the straight-forward extension to the conditional VaR and ES is presented.

Chapter 4 studies the conditional VaR in a general class of volatility models. A popular two-step estimator is examined and a fixed-design residual bootstrap method is proposed to mimic its finite sample distribution. Theoretical results

are established that underpin the method's validity and bootstrap intervals are constructed for the conditional VaR. A simulation study reveals that the common equal-tailed percentile bootstrap interval tends to fall short of its nominal value, whereas the reversed-tails bootstrap interval yields accurate coverage. Compared to the related recursive-design bootstrap, both methods perform equally well regarding coverage, yet the fixed-design scheme leads on average to shorter intervals. The fixed-design residual bootstrap method is easy to implement and illustrated in an empirical application to French stock market data.

Chapter 5 revisits the general class of volatility models of Chapter 4 and focuses on the conditional Expected Shortfall. Again, a two-step estimator is considered and a fixed-design residual bootstrap method is employed to resemble its finite sample distribution. The asymptotic validity of the bootstrap scheme is established and a simulation study confirms that the method performs adequately in samples of modest size.

While this thesis contributes to the understanding of constructing bootstrap intervals for conditional risk measures, it is far from complete and several avenues of research are left unexplored. Subsequently, we point out possible extensions of this thesis. A natural continuation would be to include a time varying conditional mean in the class of volatility models, i.e. $\epsilon_t = \mu_t + \sigma_t \eta_t$, where both, μ_t and σ_t , are measurable function of past observations: $\mu_{t+1} = \mu(\epsilon_t, \epsilon_{t-1}, \dots; \theta_0)$ and $\sigma_{t+1} = \sigma(\epsilon_t, \epsilon_{t-1}, \dots; \theta_0)$. Such framework would nest ARMA-GARCH models, for which estimation results (cf. Francq and Zakoïan, 2004) and some bootstrap results (Shimizu, 2010) are available. Whether these bootstrap methods extend to conditional risk measures remains an open issue.

In addition, a multivariate extension of the results is worth to consider. Good progress is done in that field by Francq and Zakoïan (2018), who study the conditional VaR of a portfolio when the individual returns follow a semi-parametric multivariate dynamic model. They also propose a recursive bootstrap scheme, yet leave its asymptotic validity for future research. It would also be interesting to compare their recursive-design bootstrap method with its fixed-design counterpart and study different interval types. Concerning the bootstrap consistency, it however seems to be easier to show the asymptotic validity for the fixed-design bootstrap method (Cavaliere et al., 2018).

Also, the fixed-design residual procedure can be employed to mimic the finite sample distribution of alternative estimators such as the one-step estimator asso-

ciated with the conditional VaR or ES (Francq and Zakoïan, 2015) or estimators based on extreme value theory (Martins-Filho et al., 2018). Whereas the former estimator may perform better than the two-step procedure in cases of heavy-tailed innovation distributions, the latter is appealing when exploring the conditional VaR and ES sufficiently far in the distribution's tail.

Another interesting path of research would be to study whether the bootstrap procedures proposed in this thesis achieve asymptotic refinements. This could also shed light on the difference between the fixed-design and the recursive-design residual bootstrap. In the literature only few asymptotic refinements results are available for GARCH(1,1) processes. Whereas Corradi and Iglesias (2008) derive higher order asymptotic refinements of the block bootstrap, Jeong (2018) establishes the second-order refinement of a recursive-design residual bootstrap. A reason for this deficit is that the Edgeworth expansions are extremely difficult to obtain for GARCH-type processes.

To conclude, *bootstrap inference for conditional risk measures* has practical relevance for banks, asset managers, pension funds, but also for regulatory institutions. This thesis proposes bootstrap methods to quantify the estimation uncertainty, provides theoretical results underpinning their validity and shows their good finite sample performance. Various extensions arise in this context, which are left unexplored ... for now.

Chapter 7

Valorisation

“Those who ignore statistics are condemned to reinvent it.”

-Bradley Efron (1938-...)

Some readers may be surprised by the fact that the term *valorization* goes back to Karl Marx and the idea of self-expansion of capital, i.e. the growth of capital through value-forming activity. On the contrary, Adam Smith, the father of modern capitalism, coined the term *human capital* to describe the workers skill set to perform labor. Merging both lines of thought, I discuss in this addendum the increase in the value of human capital (knowledge) through value-forming activity (my research). In particular, I stress the relevance of the topic *bootstrap inference for conditional risk measures* and highlight the innovativeness of this dissertation as well as the practical implementability of its methods.

To recognize the relevance of the topic, I explore the three fundamental pillars on which this dissertation is grounded: (i) conditional volatility models, (ii) bootstrap methods and (iii) risk measures. Emphasizing the importance of the first pillar, Robert Engle was awarded the Nobel Price in Economics in 2003 for his methods of analyzing time series with time-varying volatility. His ARCH model specifies the swing between turbulent and calmer periods in stock markets and its GARCH (1,1) extension is known as the workhorse of financial application. The second pillar relates to the revolutionary idea of using resampling techniques to perform statistical inference. In recognition of the bootstrap, Bradley Efron will accept the International Prize in Statistics in 2019, possibly the highest honor in this field.¹ Present-day bootstrap methods are the conventional solution in practice to construct confidence intervals that address estimation uncertainty. The third pillar of this thesis deals with the assessment of risk attributed to financial assets. Risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES) are extremely popular in Finance to quantify the exposure to risk, which is essential for making informed decisions.

The financial econometric thesis at hand is the result of pooling these fundamental insights from finance, economics and statistics. It proposes to construct bootstrap intervals for conditional VaR and ES that account for estimation uncertainty. Existing work shows that simulation results are promising, however there are no theoretical results underpinning the validity of these methods. This dissertation aims to fill this gap in the literature by providing the necessary theoretical foundations. In particular, the contribution of this dissertation to the literature is threefold. First, a fundamental issue is solved that arises in the analysis of econometric forecasting techniques. Second, refined bootstrap intervals are proposed

¹In 2005 he was already awarded with the National Medal of Science. If there existed a Nobel Price in Statistics, Efron would have received it *almost surely*.

for conditional VaR and ES and theoretical results (based on mild assumptions) are established confirming their validity. Proving the validity of the methods is crucial as an application of invalid bootstrap methods for predicting risk can have severe impact. Third, simulations are presented supporting the theoretical results and practical recommendations are made to practitioners.

The last point immediately gives rise to the practical implementability of the bootstrap methods. Contrary to techniques based on asymptotic theory, the proposed bootstrap methods are easy-to-implement and fully data-driven as illustrated in an empirical exercise (see p. 120). In addition, the recommended fixed-design residual bootstrap technique is compatible with various conditional volatility models (e.g. GARCH, T-GARCH, etc.) and moreover faster than its recursive-design competitor. Furthermore, the significant increase in the availability of computing power makes the computational cost of the bootstrap a subordinate concern. Nevertheless, the thesis also offers variants that find a compromise in the trade-off between computational time and precision (see p. 104).

All in all, this dissertation offers valuable insights for financial institutions that are engaged in risk management such as banks, pension funds and insurance companies. Besides their intrinsic motivation to mitigate risk to an acceptable level, they are also committed by law to calculate capital requirements by means of VaR and ES and to report them to the regulatory authorities.² Therefore, the content of this dissertation also has a immediate appeal to regulators that are primarily concerned about excessive risk-taking. I sincerely hope that the methods presented within this thesis become part of financial practice in the future and that they find their way into the regulatory frameworks to ensure financial stability.

²In the banking sector the regulatory framework is called Basel III, whereas its counterpart in the insurance sector is known as Solvency II.

Bibliography

“Wisdom is not a product of schooling but of the lifelong attempt to acquire it.”
-Albert Einstein (1879-1955)

- Acerbi, C. and D. Tasche (2002a). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes* 31, 379–388.
- Acerbi, C. and D. Tasche (2002b). On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 487–1503.
- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21(1), 243–247.
- Andrews, D. W. K. (1983). First order autoregressive processes and strong mixing. *Cowles Foundation Discussion Papers* 664, 1–23.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9(3), 203–228.
- Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics* 37(3), 577–580.
- Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74(1), 3–30.
- Bao, Y. (2018). The asymptotic covariance matrix of the QMLE in ARMA models. *Econometric Reviews* 37(4), 309–324.
- Barndorff-Nielsen, O. E. and D. R. Cox (1996). Prediction and asymptotics. *Bernoulli* 2(4), 319–340.
- Belyaev, Y. and S. Sjöstedt-De Luna (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability* 37(3), 807–822.
- Berkes, I. and L. Horváth (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* 105(2), 271–298.
- Beutner, E., A. Heinemann, and S. Smeekees (2018). A residual bootstrap for conditional value-at-risk. *Preprint arXiv:1808.09125*.
- Beutner, E., A. Heinemann, and S. Smeekees (2019a). A general framework for prediction in time series models.

- Beutner, E., A. Heinemann, and S. Smeekees (2019b). A justification of conditional confidence intervals. *Preprint arXiv:1710.00643*.
- Billingsley, P. (1986). *Probability and Measure* (2nd ed.). New York: John Wiley & Sons.
- Blasques, F., S. J. Koopman, K. Lasak, and A. Lucas (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting* 32(3), 875–887.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Bollerslev, T. and H. O. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73(1), 151–184.
- Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52(1-2), 115–127.
- Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* 121(10), 2331–2360.
- Box, G. E. and G. M. Jenkins (1971). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys* 2(2), 107–144.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. New York: Springer.
- Cai, Z. and X. Wang (2008). Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics* 147, 120–130.
- Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18(01), 17–39.
- Castillo, I. and J. Rousseau (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Annals of Statistics* 43(6), 2353–2383.

- Cavaliere, G., I. Georgiev, and A. M. R. Taylor (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Reviews* 32(2), 204–219.
- Cavaliere, G., R. S. Pedersen, and A. Rahbek (2018). The fixed volatility bootstrap for a class of ARCH(q) models. *Journal of Time Series Analysis* 39, 920–941.
- Chang, Y. and J. Y. Park (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* 24(4), 379–400.
- Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* 6(1), 87–107.
- Christoffersen, P. and S. Gonçalves (2005). Estimation risk in financial risk management. *The Journal of Risk* 7(3), 1–28.
- Corradi, V. and E. M. Iglesias (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics* 144(2), 500–510.
- Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics*. Budapest: Akadémiai Kiadó.
- D’Aristotile, A., P. Diaconis, and D. Freedman (1988). On merging of probabilities. *Sankhyā: The Indian Journal of Statistics, Series A* 50(3), 363–380.
- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.
- Davydov, Y. and V. Rotar (2009). On asymptotic proximity of distributions. *Journal of Theoretical Probability* 22(1), 82–98.
- Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1), 83–106.
- Doukhan, P. (1995). *Mixing: Properties and Examples*. New York: Springer.
- Dudley, R. M. (2002). *Real analysis and Probability*. Cambridge: Cambridge University Press.
- Dufour, J.-M. and A. Taamouti (2010). Short and long run causality measures: theory and inference. *Journal of Econometrics* 154(1), 42–58.

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.
- Falk, M. and E. Kaufmann (1991). Coverage probabilities of bootstrap-confidence intervals for quantiles. *The Annals of Statistics* 19(1), 485–495.
- Francq, C., L. Horváth, and J.-M. Zakoïan (2016). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics* 14(2), 353–382.
- Francq, C. and J.-M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10(4), 605–637.
- Francq, C. and J.-M. Zakoïan (2006). Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory* 22(5), 815–834.
- Francq, C. and J.-M. Zakoïan (2011). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: John Wiley & Sons.
- Francq, C. and J.-M. Zakoïan (2012). Risk-parameter estimation in volatility models. Working paper. <https://mpira.ub.uni-muenchen.de/41713>.
- Francq, C. and J.-M. Zakoïan (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics* 184(1), 158–173.
- Francq, C. and J.-M. Zakoïan (2018). Estimation risk for the var of portfolios driven by semi-parametric multivariate models. *Journal of Econometrics* 205(2), 381–3401.
- Gao, F. and F. Song (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24(5), 1404–1424.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48(5), 1779–1801.
- Gospodinov, N. (2002). Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics* 111(1), 85–101.

- Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.
- Hall, P. and M. A. Martin (1988). On bootstrap resampling and iteration. *Biometrika* 75(4), 661–671.
- Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71(1), 285–317.
- Hamadeh, T. and J.-M. Zakoïan (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141(1), 488–507.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, B. E. (2006). Interval forecasts and parameter uncertainty. *Journal of Econometrics* 135(1), 377–398.
- Hartz, C., S. Mittnik, and M. Paolella (2006). Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics & Data Analysis* 51(4), 2295–2312.
- Heinemann, A. and S. Telg (2018). A residual bootstrap for conditional expected shortfall. *Preprint arXiv:1811.11557*.
- Hidalgo, J. and P. Zaffaroni (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics* 141(2), 835–875.
- Hjort, N. L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *Preprint arXiv:1107.3806*.
- Huber, P. J. (2009). *Robust Statistics* (2nd ed.). New York: John Wiley & Sons.
- Ing, C.-K. and C.-Z. Wei (2003). On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85(1), 130–155.
- Ing, C.-K. and C.-Z. Wei (2005). Order selection for same-realization predictions in autoregressive processes. *The Annals of Statistics* 33(5), 2423–2474.
- Jeong, M. (2017). Residual-based GARCH bootstrap and second order asymptotic refinement. *Econometric Theory* 33(3), 779–790.

- Jeong, M. (2018). Residual-based GARCH bootstrap and second order asymptotic refinement-addendum. *Econometric Theory* 34(3), 704–704.
- Kabaila, P. (1999). The relevance property for prediction intervals. *Journal of Time Series Analysis* 20(6), 655–662.
- Kabaila, P. and Z. He (2004). The adjustment of prediction intervals to account for errors in parameter estimation. *Journal of Time Series Analysis* 25(3), 351–358.
- Kabaila, P. and K. Syuhada (2008). Improved prediction limits for AR(p) and ARCH(p) processes. *Journal of Time Series Analysis* 29(2), 213–223.
- Kabaila, P. and K. Syuhada (2010). The asymptotic efficiency of improved prediction intervals. *Statistics & Probability Letters* 80(17), 1348–1353.
- Kim, H.-J. (2008). Moments of truncated student-*t* distribution. *Journal of the Korean Statistical Society* 37(1), 81–87.
- Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association* 101(475), 980–990.
- Kreiss, J.-P. (2016). Discussion: bootstrap prediction intervals for linear, non-linear and nonparametric autoregressions. *Journal of Statistical Planning and Inference* 177, 28–30.
- Kunitomo, N. and T. Yamamoto (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* 80(392), 941–950.
- Lewis, R. and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16(3), 393–411.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- Mandelbrot, B. (1963). A long memory property of stock market returns and a new model. *Journal of Business* 36(4), 394–419.
- Martins-Filho, C., F. Yao, and M. Torero (2018). Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* 34(1), 23–67.

- Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications* 29(2), 309–315.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6(3), 318–334.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2), 347–370.
- Osmundsen, K. K. (2018). Using expected shortfall for credit risk regulation. *Journal of International Financial Markets, Institutions and Money* 57, 80–93.
- Pan, L. and D. N. Politis (2016a). Bootstrap prediction intervals for linear, non-linear and nonparametric autoregressions. *Journal of Statistical Planning and Inference* 177, 1–27.
- Pan, L. and D. N. Politis (2016b). Bootstrap prediction intervals for Markov processes. *Computational Statistics & Data Analysis* 100, 467–494.
- Pascual, L., J. Romo, and E. Ruiz (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis* 25(4), 449–465.
- Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50(9), 2293–2312.
- Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford: Oxford University Press.
- Phillips, P. C. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics* 9(3), 241–261.
- Roussas, G. G. (1997). *A Course in Mathematical Statistics* (2nd ed.). San Diego: Academic Press.
- Samaranayake, V. A. and D. P. Hasza (1988). Properties of predictors for multivariate autoregressive models with estimated parameters. *Journal of Time Series Analysis* 9(4), 361–383.
- Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies* 62(4), 639–661.

- Shimizu, K. (2010). *Bootstrapping Stationary ARMA-GARCH models* (1st ed.). Wiesbaden: Vieweg+Teubner.
- Spierdijk, L. (2016). Confidence intervals for ARMA-GARCH value-at-risk: the case of heavy tails and skewness. *Computational Statistics & Data Analysis* 100, 545–559.
- van der Vaart, A. W. (2000). *Asymptotic Statistics* (1st ed.). Cambridge: Cambridge University Press.
- Vidoni, P. (2004). Improved prediction intervals for stochastic process models. *Journal of Time Series Analysis* 25(1), 137–154.
- Vidoni, P. (2009a). Improved prediction intervals and distribution functions. *Scandinavian Journal of Statistics* 36(4), 735–748.
- Vidoni, P. (2009b). A simple procedure for computing improved prediction intervals for autoregressive models. *Journal of Time Series Analysis* 30(6), 577–590.
- Vidoni, P. (2017). Improved multivariate prediction regions for Markov process models. *Statistical Methods & Applications* 26(1), 1–18.
- Xiong, S. and G. Li (2008). Some results on the convergence of conditional distributions. *Statistics & Probability Letters* 78(18), 3249–3253.
- Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18(5), 931–955.

Nederlandse Samenvatting

*“Economics is like the Dutch language:
Im told it makes sense, but I have my doubts.”
-John Oliver (1977-...)*

Risicomaatstaven spelen een belangrijke rol in financieel risicobeheer en worden in de huidige regelgeving toegepast om de financiële stabiliteit te waarborgen. Vooral Value-at-Risk en Expected Shortfall, kortweg VaR en ES, zijn populaire risicomaatstaven om kapitaalreserves te berekenen en de marktrisico's van financiële activa te beoordelen. Risicomaatstaven nemen vaak de tijdafhankelijkheid van financiële gegevens in acht. Dit betekent dat ze *voorwaardelijk* worden toegepast, m.a.w. afhankelijkheid van verleden waarden wordt expliciet gemodelleerd. Een veelvuldig gebruikte methode om onzekerheid rond de puntschattingen van risicomaatstaven te kwantificeren is het opstellen van betrouwbaarheidsintervallen door middel van de bootstrap.³

Dit proefschrift levert drie belangrijke bijdrages. Ten eerste biedt het een theoretische onderbouwing voor veel voorkomende geconstrueerde intervallen rond puntschattingen van voorwaardelijke objecten, zoals voorwaardelijke risicomaatstaven. Ten tweede, nieuwe bootstrap methoden worden voorgesteld om de onzekerheid rond de puntschattingen van de voorwaardelijke VaR en ES te kwantificeren. Bijbehorende theoretische resultaten worden gepresenteerd ter bevestiging van hun validiteit. Ten derde, simulatieresultaten evalueren de prestaties van de voorgestelde methodiek en bijbehorende aanbevelingen zijn gericht op onderzoekers die kiezen voor een praktische toepassing van deze bootstrap methoden.

Hoofdstuk 2 beschouwt het kwantificeren van de parameteronzekerheid rond puntschattingen van voorwaardelijke objecten. Er is sprake van een fundamenteel probleem. Enerzijds wordt de steekproef als gerealiseerd beschouwt zodat het verleden informeert over het heden. Anderzijds wordt de steekproef ook gezien als een reeks toevalsvariabelen zodat men schattingsonzekerheden kan verklaren. Om dit probleem te omzeilen, worden betrouwbaarheidsintervallen opgesteld onder de onrealistische aanname twee onafhankelijke processen te observeren. Een alternatieve, realistische rechtvaardiging voor dit type intervallen wordt gepresenteerd op basis van een steekproef-splitsing en een zwakke afhankelijkheidsconditie. Omdat de voorwaardelijke objecten variëren door de tijd heen, wordt een *merging* concept gebruikt dat zwaake convergentie generaliseert.

Hoofdstuk 3 belicht de algemene opzet van het vorige hoofdstuk. Een technisch rapport verifieert de voorwaarden van hoofdstuk 2 voor verschillende tijdreeksmodellen. Deze omvatten onder meer de voorwaardelijke verwachting in een AR(1)

³*Bootstrap* betekent letterlijk laarzenriem en is afgeleid van de Engelse uitdrukking pull oneself up by one's bootstraps.

en ARMA(1,1) model, de voorwaardelijke variantie in een GARCH(1,1) model evenals de voorwaardelijke volatiliteit in een T-GARCH model. Ook wordt de uitbreiding van de voorwaardelijke VaR en ES gepresenteerd.

In hoofdstuk 4 wordt de voorwaardelijke VaR besproken in een algemene categorie van volatiliteitsmodellen. Een populaire twee-staps schatter is onderzocht en een *fixed-design residual bootstrap* methode wordt voorgesteld om de onbekende verdeling van de oorspronkelijke steekproef te benaderen. Theoretische resultaten ondersteunen de validiteit van deze bootstrap methode en bootstrap intervallen worden opgesteld voor de voorwaardelijke VaR. Een simulatiestudie toont aan dat het gebruikelijke *equal-tailed percentile* bootstrap interval de tendens heeft te kort te schieten ten opzichte van de nominale waarde, terwijl het *reversed-tails* bootstrap interval een accurate dekking oplevert. In vergelijking met de gerelateerde *recursive-design* residual bootstrap dekken beide methoden even goed, maar de *fixed-design* opzet leidt gemiddeld tot kortere intervallen. De *fixed-design residual bootstrap* methode is eenvoudig te implementeren en wordt geïllustreerd in een empirische toepassing op Franse beursgegevens.

In hoofdstuk 5 ligt de focus op voorwaardelijke ES als risicomaatstaf, waarbij dezelfde algemene categorie volatiliteitsmodellen in acht wordt genomen als in hoofdstuk 4. Wederom is een twee-staps schatter bestudeerd en een *fixed-design residual bootstrap* methode gebruikt om de steekproefverdeling te benaderen. De asymptotische validiteit van de bootstrap algoritme wordt theoretisch bewezen en een simulatiestudie bevestigt dat de methode adequaat presteert in de meeste gevallen.

De conclusie van het proefschrift is te vinden in hoofdstuk 6. De algemene conclusie luidt dat bootstrap inferentie voor voorwaardelijke risicomaten praktische relevantie heeft voor banken, vermogensbeheerders en pensioenfondsen, maar ook voor regelgevende instituten. De theoretische resultaten uit dit proefschrift ondersteunen onderzoekers om de bootstrap methoden zorgvuldig toe te passen.

Curriculum Vitae

“All grown-ups were once children . . . but only few of them remember it.”
-Antoine de Saint-Exupry (1900-1944)

Alexander Heinemann was born on August 1, 1989 in Bonn, Germany. He attended high school between 2000 and 2009 at Städtisches Gymnasium in Rheinbach while spending one year at Franklin Central High School, Indianapolis, United States. After receiving the Abitur and his civil service in Ecuador, he studied Econometrics and Operations Research at Maastricht University. During his undergraduate studies, he visited University of São Paulo for one semester and worked part-time at the KfW IPEX-Bank. In August 2013 he obtained his B.Sc. degree with a specialization in Econometrics with distinction (cum laude). Subsequently, he studied at Centro de Estudios Monetarios y Financieros in Madrid, Spain, and received his M.Sc. in Economics and Finance with distinction (Luis Angel Rojo Prize). After graduation, Alex joined the Department of Quantitative Economics as a Ph.D. candidate in September 2015, under the supervision of Dr. Eric Beutner and Dr. Stephan Smeekes. He spent two months at the Centre de Recherche en Economie et Statistique (CREST) in Paris as a guest researcher. The findings of this research are presented in this dissertation. Alexander presented his work at various international conferences, such as the Annual Symposium of the Spanish Economic Association in Girona, the Annual Netherlands Econometric Study Group Meeting (NESG) in Leuven and Amsterdam, the International Conference on Computational and Financial Econometrics (CFE) in London and Pisa, the International Conference on Econometrics and Statistics in Hong Kong, the German Probability and Statistics Days in Freiburg and the Workshop on Financial Econometrics in Örebro. Chapters of this book are currently under revision in international refereed academic journals.



